Non-Technical Aspects of Quantitative and Data Work

13

13.1 First Principles 597
Framework 598
The "Multiple I" Approach 600
Roles and Responsibilities 601
Analysis Cheatsheet 603
13.2 Project Life Cycle 603
Marketing 604
Initial Contact 606
Client Meetings 606
Assembling the Team 607
Team Meetings 608
Proposal 609
Contracting and IP 612
Project Planning 613
Information Gathering . . . 615
Quantitative Analysis 616
Interpreting the Results . . 616
Reporting and Deliverables 616
Invoicing 618
Closing the File 618
13.3 Lessons Learned 619
About Clients 619
About Consultants 623
13.4 Business Development . . . 626
Basics 626
Clients and Choices 626
Building Trust 627
Improving Trust 630
13.5 Technical Writing 632
Basics 632
Components 634
Traits 638
Examples 642
13.6 A Conversation With 643
13.7 Exercises 649

by **Patrick Boily**, with contributions from **Jen Schellinck**, **Bronwyn Rayfield**, and **Colin Daniel**

With solid analytical and abstraction skills, individuals with a background in mathematics and statistics are in high demand. The gap between theory (or textbook applications) and real-world uses can prove surprisingly difficult to navigate, however, and can lead to challenges for first-time practitioners.

In this chapter, learners are introduced to various crucial non-technical aspects of quantitative and/or data work.

13.1 First Principles

The key component of data analysis and quantitative consulting is the ability to apply **quantitative methods** to business problems to obtain **actionable insight.** But it is impossible for any given individual to have expertise in **every** field of mathematics, statistics, and computer science. In our experience, the best consulting output is achieved when a small team of consultants possesses **expertise** in 2 or 3 areas, a **decent understanding** of related disciplines, and a **passing knowledge** in a variety of other domains.

This includes **keeping up with trends**, implementing **knowledge redundancies** on the team, being **conversant in non-expertise areas**, and **knowing where to find information** (online, in books, or external resources).

We present an overview of a variety of "domains" related to **quantitative analysis** in other chapters:

- survey sampling and data collection
- data processing
- data visualization
- statistical methods
- queueing models
- machine learning
- simulations
- optimization
- Bayesian data analysis
- anomaly detection and outlier analysis
- feature selection and dimensions reduction
- trend extraction and forecasting
- etc.

The domains are not free of overlaps. Large swaths of **data science** and **time series analysis** methods are quite simply **statistical** in nature, for instance, and it is not unusual to view **optimization** and **queueing** methods as sub-disciplines of **operations research**.

By design, our treatment of these topics will be **brief** and **incomplete**. Each chapter is directed at learners who have a background in quantitative methods, but not necessarily in the topic under consideration.

Our goal is to provide a quick "**reference map**" of the topic, together with a general idea of common **challenges** and **traps**, to highlight opportunities for application in a consulting context.

These chapters are not always meant to be comprehensive surveys: they often focus solely on **basics** and **talking points**. More importantly, a copious number of references are also provided.

We will complement some of these topics with write-ups of real-world consulting projects. For the time being, however, we focus on the **non-technical aspects of quantitative work**. Note that these are not just bells and whistles; analysts that neglect them will see their projects fail, no matter how cleverly their analyses were conducted.

This chapter is a companion piece to Chapter 14 (*Data Science Basics*); the latter contains a fair amount of must-read material for would-be data scientists and consultants, including:

- objects, attributes, and datasets
- modeling strategies and information gathering
- ethics in the data science context
- the "analytical" workflow
- roles and responsibilities of data analysis teams
- asking the right questions

In the rest of this section the terms **consultants**, **data scientists**, and **data analysts** are used interchangeably, as are the terms **clients** and **stakeholders**.[1]

1: There are, to be sure, important differences: quantitative consultants do not have to be data people, and the relationship between employers/stakeholders and employees (a position held by quite a few data scientists) is of a distinct nature than that between client and consultant, but there are enough similarities for the analogy to be useful. Failing that, it could be a clever idea for data analysts and data scientists to get a sense for what motivates the consultants that might be brought in by their employers.

13.1.1 The Consulting/Analysis Framework

> The perfect consultant/data scientist is both reliable and extremely skilled; in a pinch, it's much better to be merely good and reliable than great but flaky. [Bronwyn Rayfield]

Consulting is the practice of providing **expertise** to an individual or organization in exchange for a **fee**.

Consultants may be hired to **supplement** existing staff (importantly, they are **NOT** hired as employees – consultants enjoy an **at-arm's-length** relationship with their client) or to provide an **external perspective**. Consulting duties could include some of the following:

- **making recommendations** to improve products or services
- **implementing** solutions
- breathing **new life** into a failing project
- **training** employees
- **re-organizing** a company's structure to remove inefficiencies, etc.

This seems straightforward, but there could be **complications**:

- Even though consultants are brought in by the organization, their presence is not always appreciated by employees. It is not too difficult to imagine how an outsider coming in and making recommendations to improve products and services, or to remove inefficiencies could be seen, in effect, as criticizing the current processes, let alone as potentially threatening employees' livelihoods, causing a fair amount of friction and pushback.
- If a consultant is brought in to implement solutions, the first question to come to mind should be: "why isn't the company implementing the solution(s) themselves?" Is it because of a lack of resource? Are there political implications?
- The same goes for breathing new life into a failing project: why is the project failing? Is it a failure of leadership or of planning? Is the project infeasible? Are they looking for a scapegoat?
- In the training scenario, consultants need to recognize exactly how much can be done in the allotted time.[2] Is the company hoping to offer the "illusion" of training? What kind of abilities the prospective trainees have? If they have the "right stuff", why are they not training themselves? If they do not have the right skill sets and cannot be trained, what consequences might that have on success and/or reputation?

2: Typically, the available time is quite short.

Consultants fall in one (or more) of the following types [80]:

Strategy Consultants focus on corporate strategy, economic policy, government policy, and so on; the projects they typically conduct for senior managers have more of an advisory nature than in implementation one;

Operations Consultants focus on improving the performance of a company's or a department's operations; they typically work with both strategy and technology people (in sales, marketing, production, finance, HR, logistics, etc.), on projects that run the gamut from advisory to implementation;

Human Resources Consultants focus on matters pertaining to human resources or on the workplace culture;

Management (Business) Consultants focus on variety of organizational concerns (this is a catch-all term to describe strategic, operational, and HR consultants);

Financial and Analytical Advisory Consultants focus on financial or analytical matters; for these consultants, subject matter expertise (tax law, risk analysis, statistics, etc.) is paramount;

Information Technology Consultants focus on development and application of IT, data analytics, security, and so on; they typical work on project, not on business-as-usual activities;

Specialized (Expert) Consultants are usually brought in for a specific task, which requires pointed expertise in a specific field.

For the purpose of this chapter, when we refer to **quantitative consultants** and/or **data scientists**, we usually mean someone who falls in one the last three categories, in short someone with expertise in a quantitative, analytical, technological, and/or technical field.

According to *International Management Consulting*, consultants benefit from:

- business **understanding** and external **awareness** (the so-called PESTLEE framework: political, economical, social, technological, legal, environment, ethics)
- being able to **manage** client relationships
- implementing the EDDD consulting process (engage, develop, deliver, disengage)
- being familiar with **various consulting tools and methods** specific to their area(s) of expertise, etc.

More specifically, good data scientists and quantitative consultants are expected to:

- have **business acumen**;
- learn how to **manage projects** from inception to completion, knowing that consultants are working with various people, on various projects, and that these people are also working on various projects;
- be able to slot into various **team roles**, recognize when to take the lead and when to take a backseat, when to focus on building consensus and when to focus on getting the work done;
- seek **personal** and **professional development**, which means that learning never stops;
- always display **professionalism** (externally and internally), a standard a behaviour and skills that need to be adhered to – take ownership of failures, share the credit in successes, treat colleagues, clients, and stakeholders with respect, and demand respect for teammates, clients, and stakeholders as well;
- act in accordance with their **ethical system**;
- hone their **analytical**, **predictive**, and **creative** thinking skills;
- rely on their **emotional intelligence**, as it is not sufficient to have a high IQ and recognize stated and tacit colleagues' and clients' needs, and
- **communicate effectively** with clients, stakeholders, and colleagues, to manage projects and deliver results.

13.1.2 The "Multiple I" Approach

While technical and quantitative proficiency (or expertise) is of course **necessary** to do good quantitative work, it is not **sufficient** – optimal real-world solutions may not always be the optimal academic or analytical solutions. This can be a difficult pill to swallow for individuals that have spent their entire education on purely quantitative matters.[3]

3: It certainly was for us...

The consultants' and analysts' focus should then shift to the delivery of **useful analyses**, obtained *via* the **Multiple "I"** approach to data science:

- **intuition** – understanding the data and the analysis context;
- **initiative** – establishing an analysis plan;
- **innovation** – searching for new ways to obtain results, if required;
- **insurance** – trying more than one approach, even when the first approach worked;
- **interpretability** – providing explainable results;
- **insights** – providing actionable results;
- **integrity** – staying true to the analysis objectives and results;

Figure 13.1: A data science team in action, warts and all [Meko Deng, 2017].

- **independence** – developing self-learning and self-teaching skills;
- **interactions** – building strong analyses through (often multi-disciplinary) teamwork;
- **interest** – finding and reporting on interesting results;
- **intangibles** – putting a bit of yourself in the results and deliverables, and thinking "outside the box";
- **inquisitiveness** – not simply asking the same questions repeatedly.

Data scientists and consultants should not only heed the Multiple "I"s at the delivery stage of the process – they can inform every other stage leading up to it.

13.1.3 Roles and Responsibilities

A data analyst or a data scientist (in the **singular**) is unlikely to get meaningful results – there are simply too many moving parts to any data project.

Successful projects require **teams** of highly-skilled individuals who understand the **data**, the **context**, and the **challenges** faced by their teammates.[4]

Depending on the scope of the project, the team's *size* could vary from a few to several dozens (or more!) – it is typically easier to manage small-ish teams (with 1-4 members, say).

Our experience as consultants and data scientists has allowed us to identify the following **quantitative/data work roles**.[5]

Project Managers / Team Leads must understand the process to the point of being able to recognize whether what is being done makes

4: Many newly-minted consultants and data scientists have not had enough experience with **effective team work**, and they are likely to underestimate the challenges that usually arise from such an endeavour.

5: Note that individuals can play more than one role on a team.

sense, and to provide realistic estimates of the time and effort required to complete tasks. Team leads act as interpreters between the team and the clients/stakeholders, and advocate for the team.[6] They might not be involved with the day-to-day aspects of the projects but are responsible for the project deliverables.

6: They may also need to shield the team from clients/stakeholders.

Domain Experts / SMEs are, quite simply, authorities in a particular area or topic. Not "authority" in the sense that their word is law, but rather, in the sense that they have a comprehensive understanding of the context of the project, either from the client/stakeholder side, or from experience. SMEs can guide the data science team through the unexpected complications that arise from the disconnect between data science team and those "on-the-ground", so to speak.

Data Translators have a good grasp on the data and the data dictionary, and help SMEs transmit the underlying context to the data science team.

Data Engineers / Database Specialists work with clients and stakeholders to ensure that the data sources can be used down the line by the data science team. They may participate in the analyses, but do not necessarily specialize in esoteric methods and algorithms. Most data science activities require the transfer of some client data to the analysis team. In many instances, this can be as simple as sending a `.csv` file as an e-mail attachment. In other instances, there are numerous security and size issues that must be tackled before the team can gain access to the data.

Data Analysts are team members who clean and process data and prepare the initial data visualizations. They have a decent understanding of quantitative methods. They typically have at most one area of expertise and can be relied upon to conduct preliminary analyses.

Data Scientists are team members who work with the processed data to build sophisticated models that provide actionable insights. They have a sound understanding of algorithms and quantitative methods, and of how they can be applied to a variety of data scenarios. They typically have 2 or 3 areas of expertise and can be counted on to catch up on new material quickly.

Computer Engineers design and build computer systems and other similar devices. They are also involved in software development, which is frequently used to deploy data science solutions.

Artificial Intelligence/Machine Learning QA/QC Specialists design testing plans for solutions that implement AI/ML models; in particular, they should help the data science team determine whether the models are able to learn.

Communication Specialists are team members who can communicate the actionable insights to managers, policy analysts, decision-makers and other stake holders. They participate in the analyses, but do not necessarily specialize in esoteric methods and algorithms. They should keep on top of popular accounts of quantitative results. They are often data translators, as well.

Another complication: data science projects can be downright **stressful**. In an academic environment, the pace is significantly looser, but

- deadlines still exist (exams, assignments, theses),
- work can pile up (multiple courses, TAs, etc.)

In the workplace, there are two major differences:

- a data science project can only really receive 1 of 3 "grades": **A+** (exceeded expectations), **A-** (met expectation), or **F** (did not meet expectations);
- while project quality is crucial, so is **timeliness** – missing a deadline is just as damaging as turning in uninspired or flawed work; perfect work delivered late may cost the client a sizeable amount of money.

Sound **project management** and **scheduling** can help alleviate some of the stress related to these issues. These are the purview of project managers and team leads, as is the maintenance of the quality of **team interactions**, which can make or break a project:

- ALWAYS treat colleagues/clients with respect – that includes emails, Slack conversations, watercooler conversations, meetings, progress reports, etc.;
- keep interactions **cordial** and **friendly** – you do not have to like your teammates, but you are all pulling in the same direction;
- keep the team leader/team abreast of **developments** and **hurdles** – delays may affect the project management plan in a crucial manner (plus your colleagues might be able to offer suggestions), and
- respond to requests and emails in a timely manner (within reason, of course).

13.1.4 Analysis Cheatsheet

We will end this section with a 12-point **TL;DR** (too long; did not read) snippet that summarizes the profession. These were collected (sometimes painfully) throughout the years (see [81] for more details).

1. Business solutions are not always academic solutions.
2. The data and models do not always support the stakeholder/client's hopes, wants, and needs.
3. Timely communication is key – with the client and stakeholders, and with your team.
4. Data scientists need to be flexible (within reason), and willing and able to learn something new, quickly.
5. Not every problem calls for data science methods.
6. There are things to be learned both from good and bad experiences.
7. Manage projects and expectations.
8. Maintain a healthy work-life balance.
9. Respect the client, the project, the methods, and the team.
10. Data science is not about how smart we are; it is about how we can provide actionable insight.
11. When what the client wants cannot be done, offer alternatives.
12. "There ain't no such thing as a free lunch."

13.2 Project Life Cycle

Based on our experience, we think that there are twelve steps in the **consulting life cycle** (see The Consulting Life Cycle and [81] for details):

1. **marketing** – getting the word out;
2. **initial contact** – start discussions with prospective clients;
3. **first meeting (and meetings)** – committing to write a proposal for the client;
4. **assembling a team**;
5. **proposal and planning** – laying out what can be done for the client;
6. **contracting, insurance, IP** – if the client agrees to the proposal, this step is crucial: do not start work until this step is cleared up;
7. **information gathering** – may include data collection and cleaning, meeting with domain experts and in-house specialists to get a sense of the context;
8. **analysis** – where quantitative skills come in to play;
9. **interpretation of results** – this is what the client actually cares about;
10. **reporting, dashboarding, deployment** – there are often multiple deliverables along the way;
11. **invoicing** – required to get paid;
12. **closing the file** – conducting a post-mortem with the client and with the team and deciding what is next.

In this section, we dig a little deeper into each of the steps.

13.2.1 Marketing

Marketing is required to let **prospective** clients know that an individual or group is **in business** (as consultants), that they possess a **specific set** of qualifications, and that they are looking for **projects** on which to work.

There are numerous marketing approaches:

- word-of-mouth
- online
- event
- newsletter
- article
- content
- niche
- reverse
- etc.

Obviously, not all these methods are applicable to every consultant and to every context. The principle underwriting marketing is simple: if prospective clients do not know that consultants **exist**, the latter cannot be found.[7]

7: Marketing is analogous to dating in this manner – **you must put yourself out there**.

In a broad sense, marketing is anything and everything done by a consultant to **legitimately** "get an in" with a prospective client and to convince them to hire their services. As with dating, attempts to get in "illegitimately" are usually regarded poorly and can easily backfire.[8]

8: Exactly what constitutes illegitimate behaviour is not always easy to determine, and may vary from one client to the next, but lies and misrepresentations are big no-nos.

Large consulting firms typically have **marketing teams** or departments – that is to say, individuals who are dedicated to finding clients and projects.

In smaller firms, marketing is usually done by the consultants **themselves**. Individual consultants and sole proprietors often join up with one another

to avoid duplicating marketing efforts and to minimize associated costs. Keep in mind, however, that this requires a certain amount of **business compatibility** and **ideological alignment**.

Marketing avenues should not be viewed in a fixed manner – not only are they constantly changing with the advent of new technologies,[9] but personal preferences and the appropriateness of a given approach may change over time.

9: It is recommended that consultants **stay up-to-date** on these technologies; a principled stand against a new tech may garner support in an echo chamber, but it can also mark you as **out-of-touch** with a younger and more general audience.

And while good marketing is necessary to consulting success (whatever form this may take), it is not a sufficient condition; there is a **marketing point of diminishing returns**, after which the results are not worth the effort.

Which avenue should a consultant select? The following questions are worth asking:

- does the avenue give a consultant an "in"?
- can it convince a client to hire the consultant's services?
- what is its genuine cost (time, financially, energy)?
- what is its initial vs. ongoing investment?
- what are the risks associated with it?
- are there universal guidelines to the approach?

As alluded to previously, what works for one project, one client, one consultant may not work for another – **beware the tyranny of past success**!

Marketing Materials Beginning quantitative consultants could benefit from some of the following avenues:

- current and customizable project-based CVs
- client testimonials (letters of reference, etc.)
- portfolio (online, offline), including personal and pro bono projects (GitHub repository, etc.)
- active social media presence
- updated and functional website/landing page
- blog articles/white papers on variety of topics
- brochures and business cards
- attending conferences and networking activities
- adverts
- etc.

We shall provide additional details for a few of these.

Project-Based CVs contain two main sections:

- **Traditional CV** (contact info, skills, selected achievements, relevant experience, education and personal development, personal)
- **Relevant Project Experience** (list, role, project description, related reports and presentations)

Other items can be added, depending on the context (publications, teaching, etc.). The traditional section should be no more than four pages (note the suggested section **order**). Projects in different domains could also be used to showcase successes and breadth of knowledge (see [82] for samples).

Social media platforms include:

- LinkedIn (consider accepting invites from people you do not know, expand your network, post regularly)
- twitter (RT articles/posts of interest, with a commentary; get to know who the experts in the fields are and follow them; post regularly)
- Facebook, Instagram, TikTok, etc.

With an online presence, there might also be a need to separate your **personal** from your **professional** online identities; it is important to avoid the common pitfalls of online use (trolling, flame wars, lousy/generational spelling, etc.), and to keep up with new tools.

Blog articles can be used for:

- content aggregation
- interacting with community
- pushing content
- showcasing communication, technical skills, and interests

These are not journal articles, so **effective communication is key**.[10] Examples can be found on the *Data Action Lab* blog [83]. Remember: any (legitimate) approach is on the table if it helps get a prospective client interested in your services; consult Marketing and [81] for more details.

10: Note that if you are going to base an article off of a project, you should make sure to obtain **client permission** first.

13.2.2 Initial Contact

Once a prospective client expresses an interest (no matter how faint) in working with a consultant, whether through email, a phone call, or some other approach, the consultant should:

- immediately dedicate a **project number**, an **email folder**, and a **folder in their file structure** to the potential project;
- capture and verify **email addresses** and **phone numbers** as soon as possibly;
- respond to advances **promptly** (without seeming too desperate), and
- show interest in the project, **even** if the subject matter is not to your liking or if you are not an expert on the required methods

It is too early to turn down a client at this stage due to lack of interest and/or qualifications;[11] the **goal** remains to gather some initial information about the project and set-up a meeting to discuss it in detail. Consult Initial Contact and [81] for more details.

11: If a project must be turned down, it is much preferable to invoke **lack of availability** or to suggest **another lead** instead

13.2.3 Client Meetings

If the client agrees to a meeting and the consultant has a meeting space, the client *may* decide to come to the consultants. If so, the consultants need to make sure that a private space is available, with:

- wi-fi connectivity and plenty of electrical outlets,
- projectors and other devices,

- water/coffee/pastries/muffins, and A/C or ventilation, etc.

If the consultant does not have a meeting space on hand, they should instead secure a shared meeting space or simply offer to go to the client.[12] Furthermore, they should:

- bring a laptop computer (with battery pack or electrical cord);
- bring identification;
- go to the bathroom before you enter the client's office;
- **refrain from being late** and arrive 15 minutes early (scope the parking/busing situation), and
- dress appropriately (business, business casual?).[13]

(E-)business cards/marketing material should be available, to be traded **before** the first meeting starts. Eye contact and small talk (weather, sports, news events, etc.) should not be neglected – consultants and analysts are not solely being gauged as technical experts, but as **human beings** as well – it is a rare client that wants to work with a "robot".

Remember that the "interview" process goes both ways; the consultant is also trying to determine if they want to work with the client. Before a contract has been agreed to, everything is still only **tentative**; as such, it is crucial to get a sense for client-consultant compatibility.

In general, it is preferable to let the client take the lead in describing their **situation** and **needs**. The consultant should then:

- take notes (or have an assistant take notes);[14]
- not let misunderstandings (acronym, details specific to their industry or company, etc.) pass by unresolved – ask for **clarification**;
- **never interrupt the client** – instead, they should wait for a natural lull in the conversation to make contributions and show that the client's needs and the underlying situation are understood;
- ensure the first meeting is not about the consultant (or at least, little of it is) – **more listening, less talking**.

Clients sometimes ask consultants to provide solutions on the spot – consultants should neither acquiesce to this nor commit to a project, a price, or a timeline at the initial meeting. If pressed, they could instead say: "I'm going to bring this information back to my team and we will evaluate the project's feasibility. You will hear from us within x days/weeks."

This stage's **objectives** are to:

- get a sense of the project's feasibility and of its suitability for the consulting team, and
- gather information about data sources and quantitative requirements, as well as the client's understanding of the same.

Consult Meetings and [81] for details.

13.2.4 Assembling the Team

Will consultants and analysts be working on this project alone? With a team? What roles are needed on the team? Who is available?

There are pros and cons to both individual work and teamwork.

12: That should be the first option offered, as a courtesy.

13: At the very least, consider wearing slacks/skirt, dress shirt, belt, dress shoes.

[Author unknown] These can differ based on cultural and epochal norms. Either way, you can adjust as necessary after the first meeting, but it is preferable to err on the side of "over-dressed" to start.

14: Ask for permission before recording anything.

Figure 13.2: Finding the right team members for a data science/quantitative consulting project can be a difficult task [DeviantArt artist k-3000].

Individual Work:

- bigger share of revenues for the consultant;
- resource management easier to handle;
- no need for team meetings;
- latitude in accepting/rejecting projects;
- nobody tells anyone what to do (except for the client, perhaps);

Teamwork:

- more available resources, so project can be completed quicker (although this is only true up to a point in practice);
- more knowledge/ideas at the team's disposal;
- only one person needs to interact with client;
- managing egos and personalities can be difficult, at times;
- there is psychological strength in numbers (in theory, at least).

Our experience suggests that teams can usually achieve more, but this can come at a price: for some consultants, the level of satisfaction derived by a project might be affected by how much compromise was needed to see it through.

But the objective remains simple: assembling the team requires finding competent and pleasant people to work with. It is worth taking the time to build a team that will **work**. Consult Assembling the Team and [81] for more details.

13.2.5 Team Meetings

Once a team has been assembled, it becomes crucial to plan for **team meetings**. Nobody likes those,[15] but they are **crucial** to the project's success.

15: Nobody we have ever met, at least.

The designated **meeting lead** should prepare an agenda and is responsible for the team sticking to it; that step is needed because team meetings can easily become **time sucks**. Team members should take such meetings seriously; remember – **the project's success depends on every consultant doing their work!**

Goals:

- keep the project is on track;
- exchange vital info between teammates (changes to the work plan, new discoveries by other members, etc.);
- keep team lead abreast of progress.

Consult Meetings (Reprise) and [81] for more details.

13.2.6 Proposal

If, after deliberations with the team, the project is deemed feasible (by the consultants **and** the clients), a proposal must be presented to the client. The proposal is the **foundation** of its eventual success – it is the opening salvo in the negotiation with the client.[16]

16: The military imagery is intentional.

One of the challenges facing consultants is how to gauge the value of the services they offer – **we tend to sell ourselves short**. The proposal justifies the monetary demands to the client; it also helps limit what is known as **scope creep**.[17]

17: The client may ask for small changes, which can grow into major headache: a different font, a different chart, a different analysis, a different dashboard, a different product, etc. Opening the door to proposal modifications also opens the door to the client taking advantage of the analysis team, even if they do not intend to do so.

Consult Proposal and Project Planning (1:05-10:40) for more details.

A proposal should read as a **letter** to the prospective client. Its content may change depending on the specifics of each project, but the following sections should figure in the final document:

- Background
- Objective and Scope
- Methodology
- Milestones and Deliverables
- Schedule and Assumptions
- Resources and Costs
- Travel and Invoicing
- Appendices – Suggested Workplan; List of Former Relevant Projects and Clients; CVs and Bios, etc.

Let us take a more in-depth look at their contents.

Background:

- introduction
- state what consultants understand of the client's organization (research this)

Objective and Scope:

- state what consultants understand of the client's problem (go back to client to clarify if needed)
- delimit the tasks ("we will do such and such", "we will not attempt to do such and such") and keep in mind that the client may require options

Methodology:

- suggest a series of steps / methods that the consultants will follow – the idea is to show the client that the team has already started thinking about their problem

- add a caveat that the data will be driving what method is used

Milestones and Deliverables:

- explicitly list the important steps and deliverables that will be produced for the client (prototype, final report, weekly progress reports, dashboard, executable code, etc.)

Schedules and Assumptions:

- provide a timeline for the milestones and deliverables, assuming that an agreement is reached by a certain date, or that the data is available by a certain date, etc.
- use **relative dates** if the client has deliverables or responsibilities for the project as well (better to err on the side of caution and **deliver on time and below cost** than the other way around!)
- establish the project authority on both the client and the consultant sides

Resources and Costs:

- list the resources that will be assigned to the project, with a short justification to reassure the client that the consultants are qualified to work on their project
- list the projected cost for each option (referring to the workplan as needed), with HST info
- reassure the client that they will not need to pay for work that is not done
- state that if more work needs to be done due to a change of scope, issues with data quality, or some other client issue,[18] the consulting team will wait until approval before starting the new work (**communication with the client is crucial!**)

18: Never call it a client error!

Travel and Invoicing:

- state what traveling costs will be charged to the client and that more expensive jaunts will only be undertaken with the client's approval
- state the invoicing policy – monthly/at milestones/upon completion, etc.

Suggested Workplan:

- provide a table with tasks and steps (follow the methodology), expected time expenditure, corresponding costs, and timelines
- produce a total estimate for the project (include the tax information)

Credentials and Credibility:

- add a list of previous clients for references
- add project-based CVs and short bios of team members
- add a list of other services offered by your team

The main objectives of the proposal is to let the clients know:

- what is understood of the problem at hand;
- what the consulting/data science team can do for them, and

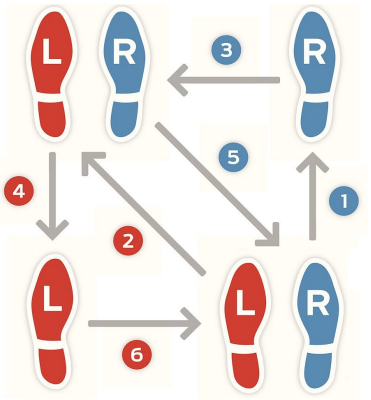
- what they are expected to do in return.

Consult [82] for samples and Proposal and Project Planning (10:50-13:25) for more details on proposals.

The Client Dance

Assuming that the proposal is sent to the client within the agreed-upon deadline you have provided them, the next step in the process is the **client dance**.[19]

19: Not usually as simple as this one:



[Getty Images]

In general, the clients **do not know** (nor do they need to know):

- how busy the consultants are;
- how many projects they have on the go, and
- how many proposals are up in the air.

Conversely, **consultants do not know**:

- the project's priority for the client;
- the procurement challenges, and
- if multiple proposals were requested from different consultants.

It is the **absence of this knowledge** that makes the client dance difficult to navigate. In the proposal, the client is given a deadline by which to respond to guarantee the availability of the consultant's resources:

- if they respond in time, then the next step is to **fine-tune** the proposal;
- if they respond after the deadline, then the consultants need to **reassess the situation** – perhaps the promised resources are not available anymore?;
- if they do not respond, the consultants must decide to either **poke them** or to **let the project go**.

If the client is not responsive, remember the **dating analogy** – sometimes the client dance is a negotiation tactic, but sometimes "they're just not that into you."

It is also important to remember that when embarking on a project with a public organization (such as the Government of Canada, for instance), the project authority on the client side is not necessarily the person who is responsible for the procurement process – that is typically conducted by a different group altogether, for whom the priorities are not necessarily in-step with the individuals the consultants have been in touch with up to that moment. All this to say that delays and complications should be expected, and are not (usually) meant as a show of disrespect.

So, from the consultants' side of things, the important questions to answer become: how desperate are they to get a project? How flexible are they? Can they afford to wait? Can they afford not to get the right terms?[20]

20: In dating terms: will they still respect themselves in the morning?

Consult Proposal and Project Planning (from 13:25 onwards), for more details on the client dance.

13.2.7 Contracting and IP

After some back and forth, the consultants and the client might agree to a proposal – this is a required step in the process, but it does not constitute a **binding legal document**.[21]

21: When analysts are working for in-company stakeholders, this section can obviously be skipped.

Consultants should **never** start work on a project until a legal contract has been signed by both parties. Some organizations insist on using their own contracts – some negotiation is possible, but they are not always willing to budge. Either way, it is crucial that consultants **REFRAIN FROM SIGNING A CONTRACT THAT THEY DO NOT UNDERSTAND OR AGREE WITH**.[22]

22: Get legal advice from actual lawyers. For real. Please.

A **contract** sets out the roles, responsibilities, and legal obligations of both parties; it contains a number of clauses, distributed in a number of sections:[23]

23: What constitutes an acceptable contract also depends on legal cultures – different jurisdictions may not use the same legal system (Ontario/Québec, for instance). We cannot over-emphasize how important it is to consult a lawyer for all your contracting questions.

- Identification of parties
- Acknowledgement of Processes by Parties
- Definitions
- Charges
- Term
- Conditions
- Payment
- Materials
- Confidential Information
- Warranties and Liability
- Force Majeure
- Termination
- Notices
- Conflict Between Documents
- Dispute Resolutions
- Waiver
- No Permanent Relationship
- Unenforceable Provisions
- Governing Law and Jurisdiction
- Singular and Plural
- Headings
- Amendment
- Language
- Fax and Counterparts
- Signatures

Contracts should never be prepared automatically – the clauses and sections may need to differ from one contract to another, although some of them may be used more frequently (see [82] for an example).

One aspect which is not explicitly listed above is that of **intellectual property**. Who owns the **results** of a consulting project? Most reasonable parties would conclude that it is the **client**; consequently, **non-disclosure agreements** (NDAs) are often required before a contract can be enacted.

But who owns the **methodology**? The **approach**? Can consultants re-use code for another project or publish the methods? Can an individual consultant use a method she developed as part of a team for her own

work? Is it even possible for mathematical, statistical, analytical work to be patented or made **proprietary**?

This might seem like a frivolous question to ask at this stage, but consultants should take the time to reach consensus on this topic with their team, and with the client – this will save everyone a lot of heartache down the road.

The goal should be to make the eventual consulting work as simple as possible by removing the focus on anything but the quantitative analysis. Consult Contracting and IP and [81] for more information.

Insurance

From the client's perspective, a consulting project only has three outcomes, of which only the first two are every (ideally) in play: either

- the consultants exceed the expectations (managed *via* the proposal and open communication), A+;
- the consultants meet their expectations, A-, or
- the consultants fail to meet their expectations, F.

Given an "F", the **best case** scenario from the consultant's perspective is that the client will simply be disappointed and send future projects to other consultants; the **worst case** scenario is that they think that the consultants also failed to meet their **contractual obligations**, opening themselves to **legal action**.

Professional insurance against this (inevitability?) is a must.[24]

24: Note that, in Canada at least, the specifics of insurance also depend on the jurisdictions in which the client and/or the consultants operate and in which the product/service is delivered. Either way, it is important to talk to a specialist **BEFORE** things go astray.

13.2.8 Project Planning

While no actual work should be started before an agreement with the client is finalized, consultants should still start **planning the project** as soon as they start working on the proposal.

The goal is to ensure that consultants meet the project's (often) tight deadline without exhausting themselves in the process, so planning will help them hit the ground running!

Note that project management is often taught as separate course in business schools and there are various lists of available references (see [84], for instance) – it is even possible to get certification (as with [85], say). Such minutia is outside the scope of this section, however.

For quantitative consultants, the most important piece of advice is to prepare a timeline for **tasks/deliverables**, incorporating:

- teammates' (and external resources') availability;
- projected delays;
- client bottlenecks;
- unexpected turns;
- holidays;
- client deadlines;
- simultaneous projects and courses;
- work-life balance, etc.,

Time Management (1 week = 168 hours)			
Students		**Consultants**	
Sleep	**56**	Sleep	**56**
Meals	**15**	Meals	**15**
Courses/study	**40**	Work	**42**
Work	**20**	Learning	**10**
Commute/errands	**15**	Commute/errands	**15**
Other	**22**	Other	**30**

Table 13.1: An example of weekly time availability for students and professionals; does it work with your work-life balance preferences? If not, what would?

while keeping *Hofstadter's Law* in mind:

> "It always takes longer than you expect, even when taking *Hofstadter's Law* into account." [86]

Consequently, consultants should be prepared to revisit their workplan periodically, especially when preparing progress reports (internal and external) – this should not be seen as a failure, but as normal and expected **course corrections**, which occur in all project work.

Weekly Schedule

It is impossible to plan the project work without having a good sense of the team members' availability (an example is provided in Figure 13.1.

There is but a finite number of hours in a week, and each of us has responsibilities outside of work, including some necessary downtime and rest.

There is no value in creating a superhuman workplan that cannot be met – consultants must be realistic if they want to deliver on time. Failure to agree to a mutually acceptable schedule means that the project **should not go forward**.[25]

25: It is infinitely preferable to realize this **before** the contract is signed; the client in under no obligation to accommodate requests for extensions after an agreement has been reached.

Workplan

If the resources are available, the first project planning step is to produce a workplan that uses **high-level phases**, with item names that follow the proposal's methodology. Once these have been nailed down (with expected durations), the next step is to break down into various **task categories and sub-tasks**, with potential deliverables, timelines, and assigned team members associated to each task.

This is as much an art as it is a science, and it can take a few sub-par projects before consultants get the hang of it. Experienced project leads can provide advice, if required, but analysts and consultants alike should be prepared to revisit and revise the workplan a number of times prior to the start of the project **AND** as the project is undertaken (see Table 13.2 for an example).

Consult [82] for samples, and Proposal and Project Planning (Reprise) for more details.

Phase 1 – Data Preparation			
Tasks	**Estimated Time (hrs)**	**Suggested Timeline**	**Deliverable**
1.1 Importing and Hosting Data	2	Nov 9 - 13, 16 - 20	
1.2 Data Inspection (i.e. data structure, metadata info)	12		
1.3 Soundness and Data Quality	51	Nov 23 - 27, Nov 30 - Dec 4, Dec 7 - 11	
1.4 Dealing with Missing Observations	37	Dec 14 - 18	
1.5 Dealing with Anomalous Observations	37	Jan 4 - 8, 11 - 15	Deliverable 1: Cleaned Data Set (Nov 9 - Jan 15)
1.6 Data Contextualization	40	Jan 18 - 22, 25 - 29	
1.7 Data Visualization and Data Description	60	Feb 1 - 5, 8 - 12, 15 - 19	
1.8 Meetings and Reporting			
1.8.1 Creating Report Describing Findings	20	Feb 22 - 26, Feb 29 - Mar 4	
1.8.2 Meeting To Review Report	1	Mar 7 - 11	Deliverable 2: Data Visualizations (Jan 18 - Mar 18)
1.8.3 Revising Report Based on Client Feedback	10	Mar 14 - 18	Deliverable 3: Data Assessment Report (Jan 18 - Mar 18)
Phase 1 – Total Estimated Time (hrs)	**270**		
Phase 1 – Total Cost @80.00/hour	**$21,600.00**		

Table 13.2: A workplan example.

13.2.9 Information Gathering

In the proposal, the consultants have demonstrated their understanding of the client's organization and of the project. But until the consultants **actually** start working on the project, that understanding may remain theoretical, at best.

Practical (and actionable) understanding can most often be gained through:

- field trips;
- interviews with subject matter experts (SMEs);
- readings;
- data exploration (even just trying to obtain the data can prove a pain),
- and other similar things.

The client is not a uniform entity – it is conceivable that (some of) its data specialists and SMEs will even **resent** the involvement of external consultants.

This stage of the process is a chance to show the various client entities that the consultants are on their side; it is also a chance to gather valuable information that was not publicly available prior to the start of the project.

This can best be achieved by:

- asking **meaningful** questions;
- taking an **honest interest** in the SMEs experiences and expertise, and
- acknowledging their ability to **help**.

This is also the consultants' first chance to identify **gaps in knowledge**,[26] which can sink a consulting project if they go undetected until they are remedied.

Much more will be said on the topic in Section 14.2.2 of *Data Science Basics* (Chapter 14) – its content should be thoroughly understood prior to embarking on this step of the process.

Consult Information Gathering and [81] for additional details.

26: **Implicit** assumptions made at various stages, either by the consultant, the client, or both. Implicit assumptions are not necessarily invalid – problems arise when they are not shared by all parties (a gap which may only reliably be discovered by attempting to gather explicit information).

13.2.10 Quantitative Analysis

If everything else has fallen correctly into place, consultants and analysts should now be itching to get going on the quantitative work.

Naturally, we assume that consultants and analysts have expertise in one or more of the following technical areas:[27]

- data collection;
- data processing;
- data visualization;
- statistical analysis;
- data science and machine learning;
- optimization;
- queueing models;
- trend analysis and forecasting;
- simulations;
- etc.

27: See [81, 87, 88] and the entirety of your degree(s) for more information ... as well as all the other chapters in this book.

This is where the bulk of the work comes in, and where **quantitative consultants** (as opposed to regular consultants) and **data scientists** get to shine.

The quantitative consultant/data analyst's job is ... well, **to get the job done**. The time for dilly-dallying is long gone.[28]

28: And not a moment too soon, if you ask us.

13.2.11 Interpreting the Results

When the analyses have been run, we obtain results. One thing to keep in mind is that clients are not actually interested in the results so much as they are interested in **insights** from analysis.

Actionable insights require results that can be interpreted and used by the client.[29] Providing the deliverables correctly is surprisingly complicated: the only way to get this fact trough to learners is through practice and experience.

29: Let it be said one last time: the best academic or theoretical solution may not be an acceptable solution in practice.

The case study write-ups available at [81] focus on the **interpretation of results** and could be used as a source of inspiration – as is often the case, none of those projects required a thorough understanding of sophisticated methods.

The objective is to figure out what is useful for the client to know about the analysis outcomes and translate the analysis results accordingly.

13.2.12 Reporting and Deliverables

If all has gone well, the project is now coming to an end.

Deliverables are **concrete** products provided by the consultants to the client in their search for **actionable** solutions. They constitute a type of **proof** that the work has been done.

They might include:

- deployment (code, software, apps), pseudo-code, conceptual ideas;

- literature review, case study write-up, recommendation(s), expert advice, popular account;
- **progress reports**, minutes of client meetings, notes, quality plan;
- **final report**, **presentation**, clean data, poster, executive summary, dashboard, user manual, white paper, technical article, etc.

Project deliverables depend on the client and on the project. Consult Reporting and Deliverables and [81] for details.

Code, software, apps should be **documented** and **tested** prior to demos and delivery.[30] Use programming guidelines and make sure that the code is devoid of unprofessional comments and variable/function names.

30: Code that does not work as it should when it should does not look very good on analysts and consultants.

Progress Reports

These let the client know what has been done, what is being done, what remains to be done:

- **keep to the essentials** – what is new, what is left to do, what SMEs are needed, timeline estimates, etc.;
- **frequency** should be arranged with the client (not more often than weekly, usually);
- can also be used **internally** (together with minutes and notes, for project management).

Final Report

The purpose of the project leads to the type of report. Typically, a final report (or potentially a storytelling dashboard, see [25]) contains at least the following sections (some can be lifted directly from the methodology):

- executive summary
- background
- objectives
- methodology
- results
- discussion / interpretation
- recommendations
- references

There will be instances where the story of the project is important (popular accounts, say), but in most instances consultants should strive to use **technical writing** (see Section 13.5). In either case, **say what needs to be said**, in a manner that is understandable and useful to the client.

The **executive summary** should include **recommendations** and **highlights** – it is directed at stakeholders and higher-ups who may not even be aware that the project has been undertaken.

Proof-read the report for spelling, grammar, and style; make the report **appealing** – forego fancy fonts and unusual font sizes. If you use mathematical symbols, consider using LaTeX or Markdown.

Samples, as always, are available at [82]. Consult Reporting and Deliverables (from 06:30 onwards) for more details.

13.2.13 Invoicing

In a sense, this step is the most important of the process: consultants cannot get paid if they do not **invoice the client**.

The invoices should be kept **simple**; the included line items should be aligned with the deliverables and the milestones described in the proposal (and/or its amendments).

Invoicing can take place

- **upfront**, such as for training sessions, say;
- at **regular intervals**, for (advisory or long-term work;
- **after** milestones and deliverables, for modular projects;
- **upon completion**, if the client is trusted, such as with a government department, or
- some **mixture** of those.

It goes without saying, the goal here is to keep money flowing to cover expenses and pay salaries.[31] Invoices should contain the consultant's contact detail, separate line items for the various deliverables, a line for the applicable taxes, payment options, and a payment deadline (30 days, 2 months, etc.), and so on.

31: Surprisingly, this is a step that some consultants have a challenging time doing – a possible explanation of this bizarre phenomenon can be found in the accompanying video.

Clients sometimes drop off the Earth without paying the invoice (an argument in favour of regular interval or milestone invoices if ever there was one); in that case, consultants need to decide for themselves when they will pursue the matter through legal means.

It is not a pleasant thought to entertain, but consultants need to be aware that this can happen in (rare) instances. It could be preferable to simply walk away without being paid, while documenting what happened (with signed contract and proof of delivery) and reporting the client to the better business bureau. In other cases, legal action could be justified.

A number of factors are at play here, so there is no one-size-fits-all approach but let us share the advice we were given when we started out as consultants: do not put all of your eggs in the same basket – only take on a project if its failure would not end your foray into consulting. It might help to know that public sector projects (such as those run by the various levels of Canadian government) are highly unlikely to turn into dine-and-dash affairs.

Consult Invoicing for more details.

13.2.14 Closing the File

The client has accepted the deliverables and has paid the invoices, and now the project is **over**. The last step in the project life cycle is the **post-mortem**, in which the team:

- analyzes the project process;
- identifies the high marks and the low points;
- plays the what-if game (how could thing have been done differently, in hindsight?);
- decides whether they would accept to take on another project with the client if the opportunity presented itself.[32]

32: There are no right or wrong answer here – remember the dating analogy: consultants have **agency**.

If they are amenable to doing so, the consulting team could also consider conducting a post-mortem with the client.[33]

The **goal** of the (internal/external) post-mortem is not to assign blame, but to **learn lessons** (see Section 13.3) that can be applied to future projects.

Consult Closing the File for additional details.

33: Fair warning: this process could be quite painful for the consultant/analyst's ego. Introspection is one thing when it is done with the team; being criticized by the client can prove quite unpleasant, even when it is not done with malice.

Every project is different, without exception. The **project life cycle** can also differ from one project to the next, or from one client to the next, but on average, most of the steps highlighted in this section will be involved in one way or another.

13.3 Lessons Learned

In the post-mortem, consultants and analysts synthesize what they learned about dealing with **clients** and **stakeholders**, as well as with their fellow **consultants** and **analysts**, throughout the project. The following lessons were (at times painfully) extracted from 40+ past projects.[34]

34: As ever, we use consultant and client interchangeably with analyst and stakeholder, respectively, but some of these lessons are only likely to be applicable in the consulting context.

13.3.1 About Clients and Stakeholders

More context for this section's content is provided in Lessons Learned: About Clients .[35]

35: Names and identifying details have been removed to preserve privacy but note the extent to which the dating analogy remains applicable.

Welcome to the Client Dance Getting a project off the ground can be exhausting (starting from when the client shows an initial interest to the time when work can start in earnest). Consultants start at a disadvantage and the client sometimes uses this as a negotiation tactic, by keeping the consultant waiting.

Solution: consultants should be polite, but they should also respect themselves, their abilities, and their work. While it will always be true that consultants need clients (to get paid), the converse is also true: clients need consultants (to get the work done).

Beware the Scope Creep and Divergent Expectations The client may start by asking for a little "something" which is not explicit in the agreement (a different font, colour scheme, etc.). These are not big demands, and the consultants may agree too do so (for several reasons).

Then the client might ask for a little something else (repeat the analysis with a slightly different dataset to reflect that new data has come in, say), and the consultant could run a few more analyses, and so on. On their own, none of these demands are "big deal", but when all the demands are added up, a whole new project has sprung up from these little bits, without the client having had to pay for it.

Solution: consultants should leave room in the agreement for modifications, with the caveat that the workplan may need to be revisited (as would timelines and costs).

It is entirely natural for the client to want something other than what they originally agreed to – as consultants start their quantitative work, they may expose conceptual and knowledge gaps, which could then lead the analysis into unexpected areas. Be that as it may, the agreement must be adhered to; modifications remain possible, but these should come at a cost to the client.

What Clients Want vs What They Need Clients do not always know what they want, from a quantitative perspective, and so what they want and what they need is not usually the same. This can come about because a previous consultant sold the client on an approach or buzzword, or because the client's competitors are doing something specific, and they feel they should follow suit.

Solution: it is the consultant's responsibility to offer advice on what the client needs, not necessarily what they want. This advice should be documented because the client might decide to disregard the consultant's advice and go with what they want (over what they need).

If the client comes to realize that what they wanted was not what they needed, the documentation should prevent the consultant turning into a scapegoat.

Talk is Cheap Some clients are very gregarious: they are full of promises and full of ideas when it comes to a project, they will engage consultants in the process, ... but for whatever reason they do not respond to the proposal, or they will not agree to a meeting, or they will not make the data available, etc.

Solution: consultants should not start work (in earnest) on a project until an official agreement has been reached.[36]

36: Some basic prep work can still be conducted, however, but not at the expense of projects that have officially been agreed to.

Disappearing Clients Some clients pull out of the process at some stage (after the initial contact, after the proposal sent, after the project has started at invoicing, etc.).

Solution: consultants should withhold deliverables until contact has been re-established with the client.[37]

37: Take the time to document attempts at reaching the client (email, phone calls, supervisors, etc.); this could come in handy at a later stage.

Consultants should learn how to sniff out disappearing or flaky clients; one way to reduce the risk is by maintaining healthy and regular communication. Consultants may have legal recourse if a client disappears at invoicing time, but they should be ready for a fight.

Helicopter Clients The opposite situation can also occur: some clients are micro-managers and want to be involved with every aspect of the project.

Solution: consultants need to learn how to sniff out helicopter clients early. The team lead should consider giving the client nominal work to do and get them out the consultants' way.[38]

s like they are
may stop in-
unreasonable

The team lead is responsible for sheltering the consulting team from this annoyance and may have to shoulder the brunt of the interactions with the client and may have to appeal to the agreement if it contains clauses that clearly delineate the responsibilities and roles of each party.

Desperate Clients In another common situation, clients sometimes turn to consultants as a final effort to save a project (or to shift the blame to an outsider). A desperate client is often identifiable by demands for unreasonable deadlines.

This toxic situation can quickly become unbearably stressful and taxing for the consultant team.

Solution: consultants should be clear about the scope, the objectives, and the deliverables, as often as required, and be ready to return to the proposal often to remind the client what has been agreed to. While consultants need to show flexibility to keep clients happy, some limits should not be crossed.

Dishonest Clients While the previous three lessons could be chalked up to clueless (and non-malicious) clients, the next one cannot. There is no sugar-coating it: some clients will knowingly try to take advantage of the consultants.[39]

39: We are not talking about miscommunication or honest mistakes, here – some clients have a track record of abusing consultants.

Solution: at times, paying work might be hard to come by, but consultants still need to do their homework and see what there is to be found about the client from external sources before an agreement is reached – it is important to trust instincts.

Sometimes, the problems only appear after the contract has been agreed to. In that case, the priorities should be for consultants to protect their team (including themselves) by document conversations and collect a (e-)paper trail.

Consultants should avoid threatening to sue the client unless they are ready to follow through with the suit; contacting a lawyer as soon as a problem arise is necessary.

Procurement Issues The proposed project authority on the client side is not always the person who holds the purse's strings, nor do they necessarily have the final say on procurement matters – they may be interested in getting the project going, but company or departmental policies could "get in the way" and complicate the process.

Most client organizations have their own internal process to hire consultants; for (small-ish) private sector clients, the issues are likely to be minimal, but for larger private sector clients and public sector clients, there are rules in place to stamp out corruption and nepotism.

Procurement vehicles include sole-source contracts, standing offers, expert advisory agreement, professional services supplying, etc.[40]

40: These vehicles require a lot of administrative set-ups on the part of the consultants, in Canada, at least [@ProServices].

In many cases with contract value thresholds (sole-source contracts, say), clients sometimes try to squeeze a large project in under a small budget, because the only reasonable alternative would constitute contract splitting, which is disallowed (in the public service, at least, you cannot give two contracts for the same project).

Solution: consultants should avoid selling themselves short, not only because it will mark them as "easy marks", but also for a more pragmatic reason: if the client can only offer $25K for a project and the consultants agree to do $50K's worth of work, the result is a $25K shortfall, which needs to be covered from somewhere else.

Speaking Truth to Power Organizations have the tendency to trust outsiders over internal experts,[41] so there could be instances when the client already has a pretty good idea of what the data is saying but they need a person who is external to the organization to relate it to stakeholders.

41: We're not sure why that is the case, to be honest – if an organization does not trust its internal experts, they are not hiring the right employees, and that is entirely on them.

And it could also be that the client knows that whatever report will be delivered will be poorly received by the higher-ups and they do not want to suffer their wrath, so the consultant is brought in as the bearer of unwelcome news.

There is nothing wrong with this, but some consultants might not enjoy being set-up to fail (as they might see it).

Solution: it is better to have all the cards on the table so that both parties know for what purpose the consulting team has been hired.

Consulting Witches and Wizards Quantitative methods are seen as mysterious for a large swath of the population; consequently, experts in the field are sometimes viewed as witches and wizards.

As practitioners of the "magic arts", consultants and data scientists are often saddled with expectations that can sadly not be met. Quantitative methods (coupled with sound data) can achieve many remarkable feats, but consultants do themselves (and their colleagues) a disservice by not managing expectations of their abilities early (and often).

Solution: consultants should be clear and direct about what they (and their methods) can do for the client and nip in the bud any delusions the client may harbour about the project's outcome.

This is also in the client's best interest and will stop them from making grandiose promises to their stakeholders – promises that simply cannot be kept.

It is always preferable to under-promise and over-deliver than *vice-versa*.

Calendars and Deadlines Time management is difficult in general, but in consulting projects, it is also difficult because clients are not always forthcoming about their own internal deadlines and calendars. They may attempt to move the project deadlines to synchronize with changes in their own deadlines.

The clients may also have deliverables for the project (getting back to the consultants in a timely manner, providing the data by a certain date, etc.).

Solution: when they hand in progress reports, consultants should remind clients of the timelines for each remaining task, as agreed to in the contract.[42]

The proposal should also reflect the effect of the client not meeting their project deadlines. Having a clause that reads "task 3 will be completed 2 weeks after the data has been delivered by the client", say, rather than "task 3 will be completed by October 10" makes it clear that if the client delivers the data on October 8, the consultants will not have to scramble to complete task 3 by October 10 – if the task takes 2 weeks to complete, the consultants should get 2 weeks to complete it.

42: There is nothing wrong with clients asking for the timeline to be revisited, and if the consultants can accommodate the new deadlines (in terms of resource availability), they should consider doing so. But the clients should not assume that a change is forthcoming just because the client's deadlines have changed.

Data Availability and Quality Invariably, the data is not as sound as the client thinks it is. And very often, due to internal politics the data will only be available way later than it was supposed to be, which can hamper the consultants' ability to complete the project on schedule.

Solution: at no point should consultants accept the client's word that the data is "good" and does not need to be cleaned/explored.[43] The proposal (and methodology) needs to reflect that data cleaning and data exploration are essential to the project's success, and that consultants cannot guarantee the work unless they have access to the data.

43: Always in a polite manner, of course.

Dealing with Adversity Even when all the stars are aligned, and the consultants did a top-notch job and the clients provided domain expertise and quality data on time, it remains possible that the analysis results will not be to the client's liking – data does not bend to anyone's wishes.

As it is possible that the consultants made errors along the way, the client may be in the right to ask for a re-do.[44] Where it becomes problematic is when they ask for a refund and/or put your credentials in doubt.

44: Validation protocols should be in place, at any rate.

Solution: the proposal should reflect the nature of quantitative projects, i.e., the data/methods do not always support the client's hopes. Consultants should not enter in a contractual agreement with clients who do not accept (and agree to) this fundamental fact.

13.3.2 About Consultants and Colleagues

More context for this section's lessons is provided in Lessons Learned: About Consultants ↗ .

Importance of the Post-Mortem It is impossible to learn any lesson if nobody knows what the lessons are. The importance of the post-mortem cannot be overstated.

Solution: post-mortems should always be conducted, even when the project was a success. It might also be a good idea to do project components post-mortem: consultants do not need to wait until the end of the project to identify what went well and what did not for a particular phase. Lessons are learned continuously.

Boom or Bust It is often the case that consultants (especially individuals and small teams) go through periods of months without a consulting project, followed by short periods where everyone wants you to work with/for them, which wreaks havoc on work-life balance.

Solution: to survive "boom" periods, consultants need stellar project management. "Bust" periods can be used for **business development** (Section 13.4), or for research and continuous learning.

Protecting Yourself Against Unreasonable Clients As discussed in the previous section, unreasonable clients will happen at some point in every consultant's career. While this might seem to be a lesson about clients, learning to deal with them is a lesson about consultants.

A project going belly-up is not the end of the world, although it can certainly seem that way in the middle of the blow-up.

Solution: consultants should get access to lawyer and a support system before the first sign of trouble – it might be too late to do so after the fact. More importantly, consultants should be something other than consultants, at times – the benefit of physical exercise, hobbies, volunteering, personal time, and so on has been demonstrated repeatedly. Do not look down your nose at these simple solutions.

Teammates as Hurdles Teamwork is not easy. Sometimes it will feel like teammates are hindrances in the pursuit of the project's success – they just do not understand what needs to be done or they focus on things that are considered superfluous or out-of-bound by other consultants.

Solution: whether the other team members are missing the boat or not, they (and by extension, the client too) need to be always treated with respect – there will be times when having them as part of the team will be an incredibly welcome development.

With rare exceptions, consulting teams achieve more than individual consultants (although that may be easier to do when well-defined consulting roles are assigned and adhered to).

Academia vs. Business World As quantitative consultants, technical skills are in high demand – they are experts and want to provide the best possible solution to their client. But their ways of thinking are often academic due to their training; they often favour the general over the specific.

In practice, theoretical solutions are not always actionable – they cannot always be turned into useful insights for the client.

Solution: consultants need to remember that it is not about them: "best results" in the consulting context means "most useful for the client", not necessarily for the consultant's publishing record – being right is important, but it is secondary to being useful.[45]

45: Most consulting work is unsuitable for publication, in our experience.

Selling Yourself Short Many quantitative people sell themselves short to try to not intimidate their contemporaries or potential clients – we would strongly suggest not doing that. Consultants have skills, and their time and work are worth something (quite often, an extremely high something).

Solution: know your worth and be confident.

Being an Arrogant Blowhard But there is a slim line between confidence and cockiness – consultants being aware of their skills and technical proficiency should not translate as them being insufferable colleagues. Sherlock Holmes, Gregory House, and Sheldon Cooper might be intriguing TV characters, but consultants who try this approach in the real world will soon find themselves deserted by colleagues and clients alike, and without work.

Solution: consultants need to remember that consulting projects are not there for them to showcase how smart they are, but to give them a chance to help their clients achieve something they could not do on their own. There are tons of qualified quantitative people in the consulting ecosystem, and more are joining the fray every year – standing out for being arrogant or a blowhard is NOT a viable marketing strategy.[46]

46: The dating analogy rears its head again: there are plenty of fish in the sea. Clients prefer their consultants to be friendly rather than annoying.

Keep Your Edge Nothing ages faster than a quantitative consultant and their expertise;[47] methods do not suddenly become invalid, but they can easily become *dépassé*.

47: When we were students, there were barely any business applications for machine learning, for instance.

Solution: consultants need to stay up-to-date and continue learning new methods and approaches on a regular basis, while maintaining their qualifications/certifications.

Keep your eyes peeled and any eventual shock will be lessened – things rarely go exactly as planned, but keeping track of these lessons and learning to recognize the warning signs can only be beneficial to quantitative endeavours in the long run.[48]

48: You may encounter some of those in your own consulting/data analysis endeavours, but you will no doubt also encounter your very own lessons over time.

13.4 Business Development

There is one aspect of the job which does not usually come naturally to quantitative people: they also must be businesspeople. **Business development** (BD) is anything that helps to develop new business (for the company or for individual quantitative workers).[49]

49: This section does not apply as obviously to employees as it does to consultants, but some of the concepts presented here can still be transferred to the former's context, insomuch as employees have control over the work they do.

Most of the material in this section comes from [89–91] and from a presentation provided by Colin Daniel (of APEX RMS); it is also covered in the following videos:

- The Basics of Business Development
- Clients and Choices
- Building Trust
- Improving Trust

13.4.1 Basics

Consultants have two types of clients: **external clients** (the usual stuff: organizations/individuals with whom they contract to offer services) and **internal "clients"**, especially in larger consulting shops, where consultants/analysts

- sell their services to project managers, and
- support project managers in delivering to their clients.

In this view, everything that a quantitative consultant does can be referred to as **providing "services" to "clients"**. BD is crucial for consultants: it allows them to be (and to remain) employed. Despite this, quantitative consultants are not usually very fond of BD, often feeling that this task is beneath them.[50]

50: This could be a gross generalization, but we cannot find any other reasonable explanation for the reticence of quantitative folks to engage in BD.

Drumming up business is not a waste of time – pragmatically, efficient BD leads to **more time** for research, development, and quantitative analysis. A better understanding of clients and their motivations is essential to design a better BD plan, which hopefully turns into **satisfied clients**,[51] which hopefully turns into **repeat clients**.

51: The analytical work still must be conducted properly, however!

And when consultants do not have to worry about where their next project is coming from, they can focus their mental energy and efforts on **offering good services**.

13.4.2 Clients and Choices

Unless consultants have also been on the client side, they may not understand what drives **client choices**.

The Client Experience

From the client's perspective, a consulting project is a **risky** endeavour. There is a level of **personal risk**: they are putting their affairs in someone else's hands and are relinquishing control over the analytical process (even if they are brought in as domain experts).

There is an axis of **insecurity**: clients may wonder whether the consultant really wants to help them or is just out to help themselves, or whether the consultant will make the problem more complex than it really is (based on past experience with consultants and/or academics). Finally, the client may be **skeptical**, having been "burned" by consultants before.

They may be concerned that the consultant will not keep them informed, will be hard to reach, or will lose interest in the problem. From the client's perspective, buying professional services is not usually a pleasant experience – they would rather be buying **solutions to their problems** rather than buying a consultant's **time**.

In an employer/employee relationship, it might seem at first glance that the opposite is true: the employee's time is in fact what is purchased by the employer. But that attitude is slowly changing for quantitative workers, especially in the post-pandemic workplace.

So how do clients choose a service provider?

The Client's Choice Process

Part of the difficulty is that qualified quantitative consultants are commonplace – unless their skills are **truly** unmatched by competitors, professionals are rarely hired because of their technical capabilities.

Excellent quantitative capabilities are required to be **considered** in the first place, but it is other things that get a consultant selected – maintaining long-term business is more about **relationships** than it is about consultants' technical proficiency. Among the set of qualified candidates, clients seek the ones they can **trust**.[52]

52: Crucially, this is a 2-way street: consultants also should be seeking clients they can trust.

13.4.3 Building Trust

Trust is a necessary (but not sufficient!) requirement to successful consulting projects.

The Trust "Equation"

Trust is built using several factors:

- credibility, reliability, and intimacy (all positive), and
- self-orientation (negative).

The relationship is sometimes expressed *via* the **trust "equation"**:

$$\text{Trust} = \frac{\text{Credibility} + \text{Reliability} + \text{Intimacy}}{\text{Self-Orientation}}.$$

- **Credibility** refers to the consultant's technical expertise and ability to project confidence in the latter in the client's mind;
- **reliability**, to dependability and consistency on the consultant's part (work done well and on time);
- **intimacy**, to the idea that business relationships require awareness of mutually increasing risk (clients and consultants are in this together), and
- **self-orientation**, to advisors who appear to be more interested in themselves than in the client.

Credibility

Consultants commonly achieve this component *via* qualifications and references, by presenting themselves in a professional manner, and by being accuracy, precise, and complete in their work.

In general, it is not obviously clear that clients can distinguish **outstanding** work from merely **competent** work, unless they are themselves experts in the field. Most clients who leave a business relationship with a quantitative consultant do not do so because of technical incompetence, but due to **small dissatisfactions with the service**.

Even sophisticated clients will come to focus on the **quality of service** rather than the quality of work. This can sometimes be baffling to recent quantitative graduates, but it is like the notion that **how we say things** matters just as much (if not more) than **what we say**, in many contexts.

Reliability

As the number of interactions with the client increases, reliability can be demonstrated with **consistent** consultant behaviour:

> on time + on spec[53] + on budget + "extra" touches.[54]

Why does this prove important? Marketing is painful and not usually that effective, in the final analysis; great customer service is probably the **most effective** and **least expensive** marketing strategy.

And this does not apply only to repeat clients; current clients can serve as references for future projects – the least they should be able to say about a consultant is that they are **reliable**.

53: Consultants providing what has been agreed upon.

54: As long as these originate with the consultant; when it is the client that asks for more, then there is the danger of scope creep.

Intimacy

According to experts, "lack of intimacy" is a common **failure in building trust**. Mutual increasing risk brings clients and consultants together; **candor** and **honesty** are crucial.

These experts also claim that consultants should aim to become clients' "friends" and confidants, but we feel less confident about that: there are power dynamics at play, and the potential for abuse exists.

We will have more to say on intimacy in the next section.

Self-Orientation

Quantitative consultants should not appear to be more interested in themselves than in their clients; self-orientation is the greatest source of client distrust. They say that the client is always right, even when they are not right... within reason.

In practice, this translates to consultants never telling the clients flat-out that **they are wrong** without offering them **alternatives** or a **way out**.[55] In the consulting world, what this really means is that the clients' **true needs** should come before the consultants' "desire to create a monument to their own technical ability". Clients will smell that something is off if the consultant is just out to pad their CV.

55: Flexibility is the consultant's ally, however: there are instances where it makes more sense for the consultant to walk away (subject to contractual obligations, of course).

The best way to appear interested in the client's project is to **BE** interested in the client's project, and that is best accomplished by taking on projects that are interesting to the consultants.

Back to Business Development

People who know a lot more about BD than we do estimate that it is "5-10 times more profitable to sell new services to an existing customer than to sell a first service to a new customer, and that it is at least 5 times as expensive to get a new client than to keep an old one" [89–91].

Let us leave the numbers aside for now: the main advantage of working with old clients is that consultants do not need to build **trust** with them, and they can move on to the quantitative part of the project sooner (while maintaining the trust, obviously).

Furthermore, "the average sale is made after the 6th contact, but the average person quits after 2nd" [89–91]. While consultants need to be able to take no for an answer,[56] they also need to realize that they do not need to make a pitch on the very first contact; contacts can be used to build trust.

56: Again with the dating analogy.

The BD order of priority should be as shown below.

	Existing clients	**New clients**
Aware of a new need	1	3
Not aware of a new need	2	4

In general, it is easier to sell to an existing client that is not yet aware of a new need than it is to sell to a new client that is aware of their needs. The same experts also claim that "at least 70% of your business should be from past clients and their referrals; [...] very profitable firms often reach 90%".

If the working relationship with the current client is great, this is where consultants should focus their efforts **first**. The key to efficient BD is to **deliver on existing projects** and then **stay connected** with the client, as future work can usually follow with lowered effort levels.

If existing clients do not need a consultant's services anymore (or in the near future), client referrals are the next option – they may know someone in their network who could use such services. Trust must still be built in these cases, but the process has been **jump-started**.

Having said all that, there is such a thing as **client fatigue**; consulting should not become a **prison**. If consultants do not like working with a client on their project (for whatever reason), they do not have to return to them indefinitely.

Most clients are reasonable and will accept a **professionally**-handled "break-up," but some will try to pressure the consultants to return against they will.[57] Consequently, consultants might benefit from having an **exit strategy**.

57: Do we even need to name the analogy?

13.4.4 Improving Trust

Serious consultants should always be seeking to **improve** the components of the Trust Equation. **Credibility** is often more important with **new clients** – presumably, repeat clients already find consultants credible. **Reliability** and **intimacy** can be improved both with old and with new clients.

Improving Credibility

Consultants can improve credibility by:

- pointing to **publications** (peer-reviewed research papers and white papers, etc.) or to **academic honours and teaching**;
- preparing **peerless marketing materials** (current and customizable project-based CVs, client testimonials, portfolio, updated and functional website, blog articles on variety of topics, (e-)brochures, business cards, social media presence, etc.), and
- **managing client expectations**: clients are satisfied when consultants deliver more than they were expecting.

Consultants should try to strike the **right balance** – prospective clients could get scared by a portfolio on steroids. Not every detail of a consultant's history needs to be publicly available; it is easy to provide clients with more information on demand. Conversely, consultants might need to spruce up their portfolio, especially early on in their career.

Improving Reliability

Consultants can improve their reliability by:

- **doing the basics** (delivering on time and on budget and solving problems instead of generating them);
- **being available** (*via* mobile and email, being proactive with status updates even when the news is not good), and
- **being responsive** (responding to questions or comments within a 24-hour window[58]).

58: Unless it has already been established that the consultant is away for a longer time period, which is allowed, of course – health and family first, always!

Consultants should also be **informing "clients"** by:

- providing **timely** budget and proactive project status updates (both internal and external – one of the biggest consulting obstacles is waiting too long to let the clients know that something is not working out);
- being **organized** (preparing for meetings and taking the lead on agenda items), and
- **managing their time** intelligently (at times, it might be preferable to turn down work – it is better for consultants to deliver an A+ to a few clients rather than a C+ or an F on many projects).

For long term clients, reliability is as important as technical competency!

Improving Intimacy

Consultants should seek opportunities to push the boundaries of the relationship, to be **candid** and offer **weaknesses** – in other words, they should avoid trying to pretend that they are **perfect**.[59]

59: There is no sugarcoat it: **perfectionism is a flaw**.

Connections can be made by both sides looking for **commonalities**, moving beyond the small talk, and **sharing personal experiences** (at user conferences, etc.). Advisors sometimes also suggest that consultants should think of clients as their friend.

This is an area where **judgement** must be applied – the potential for abuse increases when people make themselves vulnerable. Neither **safety**, **well-being**, nor **dignity** should be sacrificed for the sake of maintaining a relationship with a client.

Reducing Self-Orientation

Consultant self-orientation may be caused by various **fears**: the fear of not knowing, of not having the right answer, of not being intelligent enough, or simply, of being rejected by the client.

It could also stem from a little streak of **selfishness** and self-consciousness, or from a need to **appear** on top of things, or from a desire to **look smart**. For clients, self-oriented consultants seem to:

- relate the client's stories to themselves;
- finish the client's sentences for them;
- need to appear clever, witty, bright;
- provide only indirect answers to the client's questions;
- be unwilling to say "I don't know";
- recite their qualifications at inappropriate times.

Clients understand that consultants are usually looking for future projects – it is not necessary for consultants to be the star of the show. As the saying goes: "you have two ears and one mouth. Use them in that proportion". Consultants should be listening to the client (and letting them talk), and by demonstrate knowledge and understanding of the client's need through **good questions**.

The best way for consultants to get what they want, which is to say, to get more paying work, is to help the client with their problems. Recommending that clients consider using other service providers, as

needed, is a particularly effective way to reduce self-orientation (and tp off-load work in "boom" periods) – it could prove useful to have a list of friendly "rivals" on hand.[60]

60: We certainly do!

It is important to realize that **not every client is going to play by the rules** – a minority of clients will try to take advantage of consultants. It might sometimes feel as though "looking out for #1" is the only reasonable approach to take.

While it remains important for consultants to protect themselves, we have found, in our experience, that focusing on reducing self-orientation is more useful overall.

13.5 Technical Writing

A consulting/data science project is only as good as how the results and recommendations are communicated, no matter how clever the analysis. We do not need to turn in reports that read like *War and Peace*, say, but the writing should not hinder the conveyance of insights; technical writing may provide an acceptable path to achieve this.

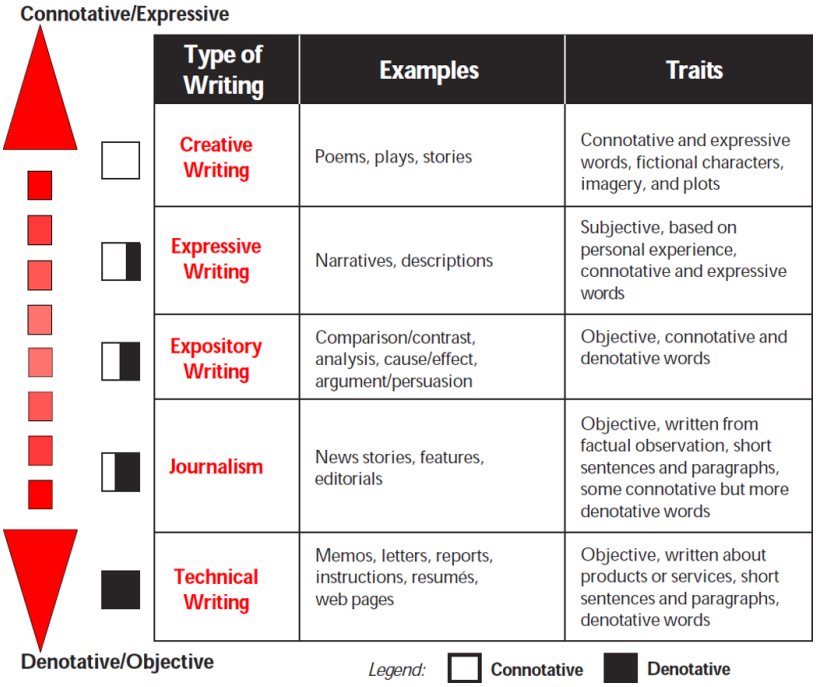
The main reference for this section is [92]; additional useful references include [93–98]. The information provided is meant to serve as a **set of guidelines**. Bend them as needed, but remain **consistent**.

13.5.1 Basics

Technical writing (TW) is communication written for and about business and industry, focusing on products and services (and policies?), and how to manufacture, market, manage, deliver, use and/or explain them. Good TW should be **precise**, **clear**, and **accurate**.

Examples of TW may include:

- CVs and résumés
- software manuals
- company websites
- instructions that come with a device
- a job description
- a falafel recipe
- help files
- code comments
- safety protocols
- official e-mails
- use cases
- case studies
- briefing notes
- research papers
- reports
- storytelling dashboards



Type of Writing	Examples	Traits
Creative Writing	Poems, plays, stories	Connotative and expressive words, fictional characters, imagery, and plots
Expressive Writing	Narratives, descriptions	Subjective, based on personal experience, connotative and expressive words
Expository Writing	Comparison/contrast, analysis, cause/effect, argument/persuasion	Objective, connotative and denotative words
Journalism	News stories, features, editorials	Objective, written from factual observation, short sentences and paragraphs, some connotative but more denotative words
Technical Writing	Memos, letters, reports, instructions, resumés, web pages	Objective, written about products or services, short sentences and paragraphs, denotative words

Figure 13.3: The communication continuum [92].

- theses
- blog articles
- etc.

TW is not prose recounting the fictional tales of characters, nor poetry which expresses deeply felt emotions through similes and metaphors; it does not narrate an occurrence/event or express an opinion; it does not report on news items; it does not focus on poetic images, nor does it describe personal experiences.

In other words, TW is neither **literature**, **journalism**, **essay writing**, nor **personal recollections**.

So what is it?

Communication Continuum

Literature is read for pleasure, essays for enlightenment, and journalism for news. TW is read to **accomplish a job**. Consequently, technical writing should be more **denotative** (provide direct definitions) than **connotative** (invoke emotional suggestions).

This **communication continuum** is illustrated in Figure 13.3.

For instance, we can compare Whitman's *When I Heard the Learn'd Astronomer* [99] (creative writing):

> *When I heard the learn'd astronomer,*
> *When the proofs, the figures, were ranged in columns before me,*
> *When I was shown the charts and diagrams, to add, divide, and measure them,*
> *When I, sitting, heard the astronomer where he lectured with much applause in the lecture-room,*
> *How soon unaccountable I became tired and sick,*

> *Till rising and gliding out I wander'd off by myself,*
> *In the mystical moist night-air, and from time to time,*
> *Look'd up in perfect silence at the stars.*

to the Wikipedia definition of **astronomy** [100] (technical writing):

> Astronomy (from the Greek: $\alpha\sigma\tau\rho o\nu o\mu\iota\alpha$, literally: the science that studies the laws of the stars) is a natural science that studies celestial objects and phenomena. It uses mathematics, physics, and chemistry in order to explain their origin and evolution. Objects of interest include planets, moons, stars, nebulae, galaxies, and comets. Relevant phenomena include supernova explosions, gamma ray bursts, quasars, blazars, pulsars, and cosmic microwave background radiation. [...]
>
> Astronomy is one of the oldest natural sciences. The early civilizations in recorded history made methodical observations of the night sky. These include the Babylonians, Greeks, Indians, Egyptians, Chinese, Maya, and many ancient Indigenous peoples of the Americas. In the past, astronomy included disciplines as diverse as astrometry, celestial navigation, observational astronomy, and the making of calendars. Nowadays, professional astronomy is often said to be the same as astrophysics.

Creative writing is "prettier" (and can be snarkier, apparently...), but technical writing conveys **precise information**. In the consulting and data science worlds, communication skills are **essential** – the best idea in the world is worthless if it cannot be communicated properly.

13.5.2 Components

TW sinks or swim based on five components (see Figure 13.4):

- **development** – preparing and presenting evidence
- **grammar** – spelling rules, syntax, conventions
- **document organization**
- **style**
- **document design** – highlighting techniques, graphs

Development

Preparing and presenting evidence should be required for all sorts of writing. TW should use examples, anecdotes, testimony, data, and research; it starts with **overall objectives**, then gets into details (items, steps, etc.), demonstrating a **logical progression** throughout.

Research often includes finding information from various sources, which should be cited when required. For quantitative writing, the **presentation of data and evidence is crucial**: TW should use paragraphs, but also charts, graphs, and tables, as necessary [92, 93].

As an example, we COULD describe how to how to put together a LEGO kit using words, but millions of kids the world over know that there is a much better approach (see Figure 13.5).

Components	Technical Writing	Essays	Summary
Development	• Uses examples, anecdotes, testimony, data, research	• Uses examples, anecdotes, testimony, data, research	Same for both
Grammar	• It is important!	• It is important!	Same for both
Organization	• Provides an introduction, body, and conclusion • Uses a subject line vs. a thesis and itemization of points vs. transitional words • Uses topic sentences only when needed, dependent upon the type and length of correspondence	• Provides an introduction, thesis statement, body paragraphs, transitional words, and topic sentences	Similar in some ways, different in others
Style	• Uses short, denotative words; short sentences; and short paragraphs	• Uses longer, connotative words; longer sentences; and longer paragraphs	Different
Document Design	• Uses **highlighting techniques,** such as graphics, headings, subheadings, various fonts, white space, bullets, etc.	• Not usually a factor	Different

Figure 13.4: The 5 components of TW; comparison with essay writing [92].

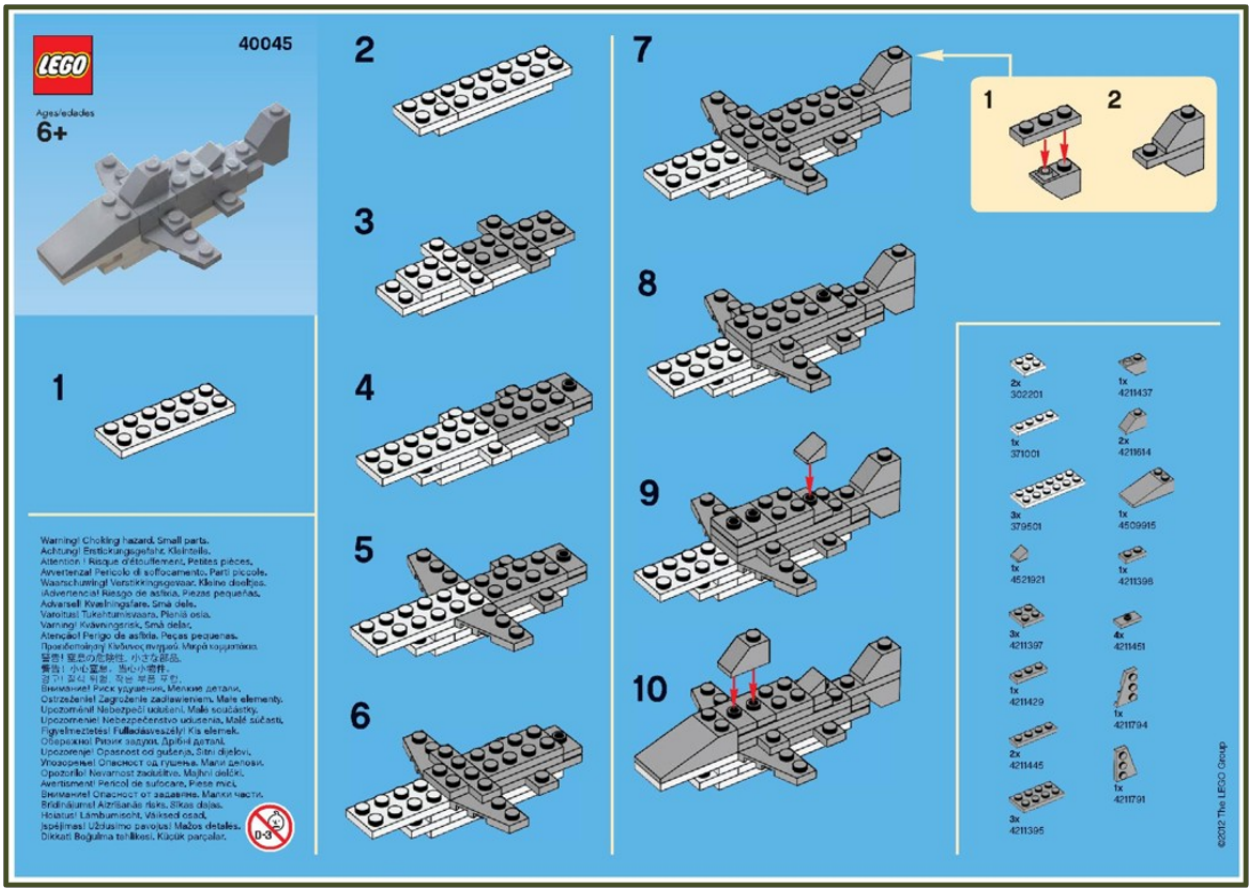


Figure 13.5: Build-a-shark LEGO instructions, with **objective** (what the assembled kit should look like), **steps** (enumerated sequence of images), and **requirements** (quantities for each piece) [LEGO].

Grammar

Consulting reports and communications which do not adhere to the common spelling and syntactic rules of English[61] (and its conventions) might not be taken seriously by some clients.

61: Obviously, the rules differ from one language to the other.

Consultants will find the following suggestions helpful:

- always use correct grammar and spelling, no matter what language their writing in – mistakes undermine what there trying to say;
- look it up or get help from someone who knows when in doubt;
- data scientists should use the second person and talk directly to the client and/or reader;
- avoid slang, dawg!;
- not be, like, real informal;
- explain acronyms: there are many possible meanings for most TLAs.

Compare with the following list of second person suggestions:

- always use **correct grammar** and **spelling**, no matter the language used – mistakes undermine what you are trying to say;
- when in doubt, **look it up** or get help from someone who knows;
- use the second person; talk directly to your client and/or reader;[62]
- avoid slang;
- do not be informal;
- explain acronyms: there are many possible meanings for most Three Letter Acronyms (TLAs).

62: Although some writers (ahem!) think that a (semi-)consistent use of person and voice is more important.

Other suggestions/guidelines include:

- using spell-checkers wisely;
- avoiding tenses shift in the middle of a sentence;
- writing sentences with a subject and a predicate (see [101] for definitions and examples);
- avoiding run-on sentences – this makes them hard to understand;
- making the antecedents of pronouns clear (see [102] for definitions and examples);
- using correct punctuation: periods ('.') end sentences and commas (',') separate dependent clauses;
- putting punctuation and footnotes inside quotation marks and parentheses (we are not a fan of this one, so we chose to ignore it);
- minimizing the use of semicolons (they tend to complicate TW);[63]
- not using apostrophes to form a plural ("Lend me your CD's!" is bad; "Lend me your CDs!" or "Lend me your CD" is better).

63: Ironic, we know.

Writing technical English is not easy, especially for those of us for whom English is not a first language. Most Canadian clients will recognize (and make allowances for) this reality, if the writing and grammar are consistent, but it remains to the consultant's long-term benefit to make an effort (or to employ an editor).

Document Organization

TW does not **usually** employ

- **topic sentences** (sentence summarizing the paragraph);
- **transitions** between and within paragraphs, and
- **thesis statements** (abstracts or summaries).

In a memo or a letter, the thesis statement is usually replaced by a **subject line**. TW uses **short paragraphs** (units of text consisting of a small number of sentences expressing a single idea, with support).

Transitional words and phrases can be replaced by:

- **enumerated lists**;
- **bullet lists**, and/or
- **headings** and **subheadings**.

TW should contain **sections**, each consisting of an **introduction**, a **body**, and a **conclusion**; the most useful, general information should go first, and it should be followed with the required details.

Style

In general, TW should use

- **short**, **denotative words**;
- **short**, **simple sentences**; and
- **short paragraphs with charts** (as required).

From a stylistic perspective, the focus should be on the **audience** and on the **purpose**.[64]

64: There are parallels with fashion and gastronomy: sometimes we need to wear a fancy suit for a special holiday meal, sometimes we need a t-shirt and jeans for a poutine.

It is important to remember that TW readers do not necessarily have an interest in the subject matter itself. Nobody reads a microwave oven's instructions for pleasure, say – TW is simply a means to an end.

Consider the following scenario: late at night on a deserted country road in the Winter, a driver hits a pothole and realizes that one of his tires has been perforated. He has never changed a tire in his entire life. Would the instructions on the Subaru website help him to do so?

> Equipment necessary to change the tire are a lift jack and a wrench. Use the jack to lift the vehicle and pick the tire up off the ground. Then use the wrench to loosen the lug nuts on the wheel. Once all the lug nuts are loose, remove them one by one and keep them in a safe place nearby. After the lug nuts are removed, the wheel and tire can be removed from the car. If a spare wheel is being put in its place, locate the spare wheel under the flooring of the trunk area, and take it out. Place the spare tire onto the lug bolts, and repeat the removal process in reverse order. Start by screwing on each lug nut, and then once all the lug nuts are screwed on, use the wrench to tighten the wheel to the disc plate. After all the lug nuts are fully tightened, disengage the jack to bring the car back to the ground. Do not exceed 50 miles per hour

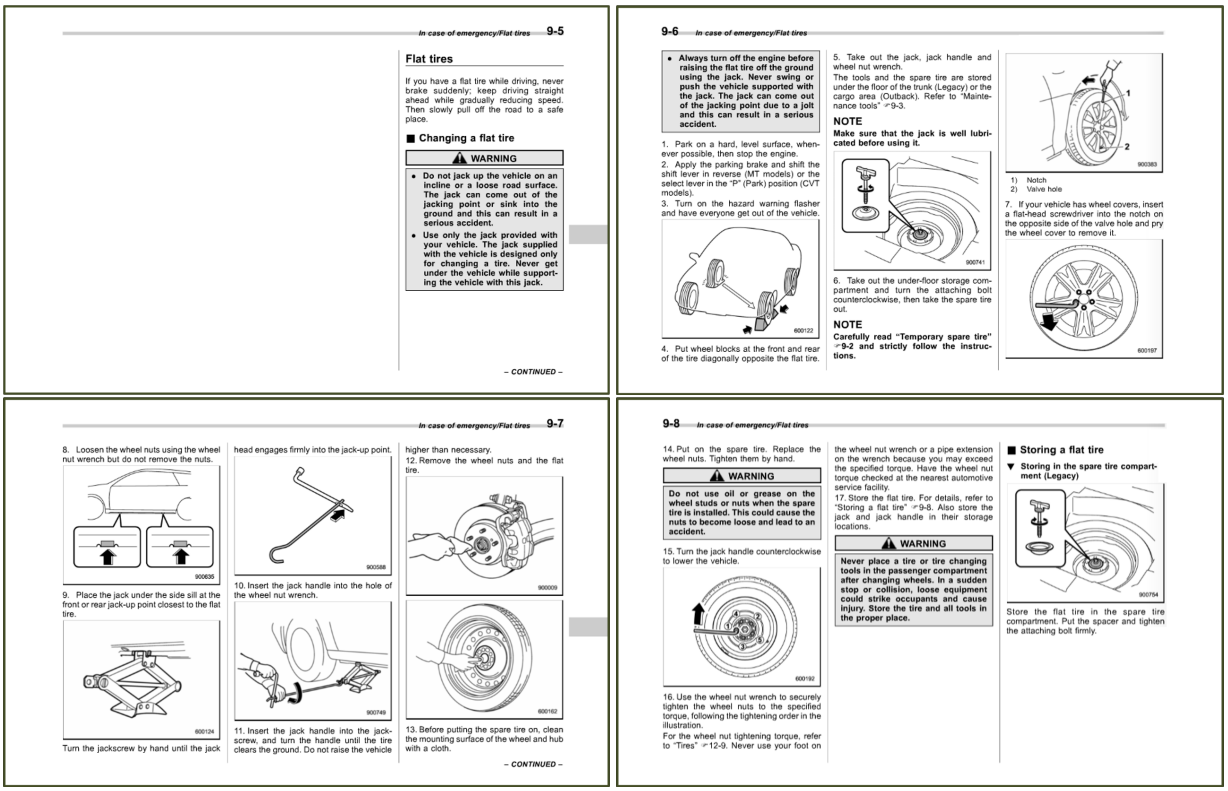


Figure 13.6: Changing a flat tire, Subaru Outback 2016 user manual [Subaru].

using a spare and changing the spare back to a standard tire as soon as possible. [Subaru.com]

Document Design

Document design refers to the **physical layout** of the correspondence. In general, TW uses **highlighting techniques**, such as graphics, lists (numbers, bullets), headings, and sectioning, but "less is more".

A small **number** of different FONTS, colours, **and** accents (**bold**, *italics*, underline) can help – **BUT** DON'T *OVERDO* it!

For sequential instructions, **numbered lists** are recommended. Longer documents could include a **table of contents** and an **index**, as well as lists of figures and tables; **hyperlinks** can be used for online documents. In all cases, clipart and low-resolution images should be **avoided**.

Let us revisit the flat tire example from above. Was the information provided precise? Did it get the message across? Was it understandable? What, if anything, is it missing? The user manual's tire changing instructions are shown in Figure 13.6. Which approach works best?

13.5.3 Traits

Sound TW exhibits five traits:

- **clarity** (organization);

- **conciseness** (fluency/choice);
- **accessible document design** (ideas and content);
- **audience recognition** (voice), and
- **accuracy** (writing conventions).

Let us discuss these traits one by one.

Clarity

The memo below is an example of unclear writing.

> From: Manager Untel
> To: New Employee Smith
> Subject: Meeting
>
> Please plan to prepare a presentation on sales. Make sure the information is detailed. Thanks.

What don't we know in this memo? What should have been included for clarity?

The **journalist's questions** (6 Ws) can help clarify communications:

- **When**'s the meeting?
- **Where**'s the meeting?
- **Who**'s the meeting for?
- **Why** is this meeting being held?
- **What** does the manager want to be conveyed about sales?
- **How** much information is "very detailed"?
- **How** will the presentation be made?

The same memo can be made much clearer, as below.

> From: Manager Untel
> To: New Employee Smith
> Subject: Sales Staff Meeting
>
> Please make a presentation on improved sales techniques for our sales staff. The meeting is planned for March 28, 2017, in Room 23, from 7:00am - 6:00pm.
>
> Our quarterly sales are down 27%. We need to help our staff accomplish the following:
> 1. Make new contacts.
> 2. Close deals more effectively.
> 3. Earn a 25% profit margin on all sales.
>
> Use the new multimedia presentation system to give your talk. With your help, I know our company can get back on track.
>
> Thanks

Clarity is the **most important criteria** for effective TW. Without it, the reader will either need to contact the writer for further clarification, or just ignore the information: the writer's and reader's time is wasted, and the message is lost.

As another example, consider a furnace maintenance safety manual. If the writing is not clear and the reader fails to understand the content, we might encounter the following consequences:

- **bad** – the furnace is damaged. The company replaces the furnace, costs accrue, and public relations have been frayed;
- **bad** – the company is sued and loses money, the writer gets fired;
- **worse** – someone gets hurt, leading to pain, anxiety, hospital bills, etc..

In a more general context, the 6 Ws should address the following items.

- **Who** is the audience? Are they beginners or experts?
- **What** do we want the audience to know or do?
- **When** will the work/event occur, in what order?
- **Where** will the work/event occur?
- **How** should the tasks be performed?
- **Why** is this information important?

It is preferable to **avoid imprecise words**, such as many, few, short, often, recently, thin, etc., and to use **precise words** and **terminology** instead, as in

> "Don't block the user interface thread for more than 2 secs."
>
> "Use four inches of 26-gauge black wire."

Another good suggestion is to **front-load** sentences with vital information, as in

> "Unfortunately, your program has timed out."
>
> "Network connection unavailable. Call 5555 for technical support."

Conciseness

Consider the 1980's referendum question:

> The Government of Québec has made public its proposal to negotiate a new agreement with the rest of Canada, based on the equality of nations; this agreement would enable Québec to acquire the exclusive power to make its laws, levy its taxes and establish relations abroad – in other words, sovereignty – and at the same time to maintain with Canada an economic association including a common currency; any change in political status resulting from these negotiations will only be implemented with popular approval through another referendum; on these terms, do you give the Government of Québec the mandate to negotiate the proposed agreement between Québec and Canada?

How easy is it to understand the question? To remember what was read? How many of you even finished reading it? How does it compare to:

> "Do you want Québec to be independent?"

Text is **concise** when it says much with few words. The idea is to keep everything short and to the point. Conciseness is important as documents must often fit in a specific physical space: a résumé being at most two pages, a car owner's manual must fit in the glove compartment, etc.

English is concise when it **avoids the passive voice**: compare "Approximately 2000 records per minute are processed by the system" (10) with "The system processes approximately 2000 records per minute" (8).[65]

65: Technical writing in French is hampered by the language's definite avoidance of conciseness.

Accessible Document Design

Consider the following paragraph:

> Regarding part number 315564-000, we received 541 units of wafer #3206-2. These were rejected. For the same part number, we received 643 units of wafer #3206-4. These were accepted. Three hundred and twenty-nine units of wafer #3206-5 from the same part number. These were accepted. Next, 344 of part number 315564-000's wafer #3206-6 were accepted. However, the 143 units of wafer #3206-7 (same part number) were rejected. Finally, all 906 units of wafer #3206-8 were rejected. These also were from part number 315564-00.

At a density of 8.4 words per sentence, the writing is **concise**; it is also **clear**, due to specificity of detail. But it is not intelligible.

Highlighting techniques open the text and make it inviting, while allowing for understanding and insight.

Document design refers to the physical layout of the communication (see previous sections). The document will be more **accessible** to the audience (in the sense that uninterested readers may still be able to digest it) if "walls of text" are avoided, and if **tables** are used to present information clearly (see Figure 13.7 for an accessible re-write of the passage above).

Audience Recognition

There are three kinds of TW audiences.

High-Tech Peers: readers in the same profession and at the same level as the writer (or higher). Example: email to counterpart in another company;

Low-Tech Peers: readers who may not have the same level of expertise as the writer but who need to understand the subject. Example: summary of a software design document written for a manager;

Lay Readers: everybody else. Example: list of a medication's side-effects, written for a patient.

TW is different for each audience type. For instance:

- high-tech peers/clients can handle acronyms and abbreviations;

Part Number 315564-00			
Wafer #	**Quantity Received**	**Accepted**	**Rejected**
3206-2	541		X
3206-4	643	☑	
3206-5	329	☑	
3206-6	344	☑	
3206-7	143		X
3206-8	906		X

Figure 13.7: Accessible description – part number 315564-000 [92].

- low-tech peers/clients might also require parenthetical definitions, and
- no acronym should be used for lay readers.

We provide an accessible description of audience recognition in Figure 13.8.

Accuracy

Finally, technical writing must be **accurate**: the information it reports must be correct and representative, with no crucial information missing. Inaccuracies can create nuisances but can also be downright dangerous.

The difference between inaccuracy and **imprecision** is illustrated by the following statements: "Use 4 feet of 3/8-inch rebar" when the requirement is for 1/2-inch rebar is **innacurate**, but "Use 4 feet of rebar" is **imprecise** as it does not specify the diameter, so the builder might not be sure.

Writers use various tricks to help with accuracy, such as:

- finishing the writing, letting it sit, then re-reading to see what might have been left out or gotten wrong;
- have someone else read it;
- reading it aloud slowly;
- reading it backwards, or upside-down,
- etc.

13.5.4 Examples

In the following pages, we provide an example of what TW could look like in a quantitative context, being a readable executive summary of the analysis of the algae blooms dataset conducted in *Data Preparation*

Achieving Audience Recognition		
Audience	**Style**	**Example**
High Tech Peers	Abbreviations/ Acronyms OK	Please review the enclosed **OP** and **EN.**
Low Tech Peers	Abbreviations/ Acronyms need parenthetical definitions.	Please review the enclosed **OP** (Operating Procedure) and **EN** (Engineering Notice).
Lay Readers	No abbreviations/ acronyms. Explanations instead.	By following the enclosed operating procedure, you can ensure that your printer will run to our engineers' desired performance levels.

Figure 13.8: Accessible description of audience recognition concepts [92].

(Section 15.7) and *Regression and Value Estimation* (Section 20.6), based on an original case study by Torgo [103]. It is obviously directed at a technical audience, but it does consider the Multiple I's as it weaves its narrative.

It is fair to say that a reasonably high number of quantitative workers got into the field to avoid writing in the first place, in the hopes that numbers and charts could carry the conversation. With an ever increasing focus on globalization, a large number of us also end up having to work in a second or even a third language, in which writing does not come naturally.[66]

66: Tools like Grammarly (syntax) and DeepL (translation) can help, although they cannot be the "be all and end all" of writing endeavours.

Be that as it may, being able to convey actionable insights to clients and stakeholders in a clear manner is fast becoming a **must** for quantitative consultants and data scientists alike.

We will give you a final tip, which is often offered to beginners: **writers must first and foremost be readers**. Read voraciously, read technical papers and *comptes rendus*, read fiction and non-fiction, read in another language. There are tons of quantitative writers on the Internet; they can provide a baseline on which to build.

13.6 A Conversation With ...

In 2020 and 2021, we recorded 10 conversations with individuals involved with quantitative consulting and data analysis, near and far.

IQC From a Student's Perspective (17:03) Smit Patel is an analyst at the Canada Border Services Agency. He is a former student in uOttawa's *Introduction to Quantitative Consulting* course. We discuss his experience as a student in the course.

Ethics in Quantitative Contexts (28:20) Julie Paquette is a professor at St-Paul University's *School of Ethics, Social Justice and Public Service*, and a Co-Director of the *Research Centre in Public Ethics and Governance*. We discuss ethics and technologies in the data context.

Multi-Tasking and Time Management (19:26) Youssouph Cissokho obtained his Ph.D. in Mathematics at the University of Ottawa in 2021. We discuss his ability to work on multiple projects simultaneously.

Consulting Experiences (26:58) Jen Schellinck is the Principal at Sysabee and an Adjunct Professor of Cognitive Science at Carleton University. We discuss her experience with consulting (2012–).

The Client's Point-of-View (24:17) Maryam Haghighi is the Bank of Canada's Director of Data Science. We discuss what consulting projects look like from the clients' perspective.

From Academia to the Workplace (19:37) Ying Gai is a senior analyst at the Canada Revenue Agency. We discuss her transition from academia to the "real world".

Proposals and Budgeting (32:28) Doug Munroe is Principal and Co-Founder of Politikos Research (P.E.I.) and a former Professor at Quest University. We discuss writing proposals and preparing budgets.

When Projects Go "Kerplunk!" (29:09) Andrew Macfie is a former freelance software developper and consultant from Toronto. He currently works as a programmer at Tucows . We discuss the ways in which consulting project can go bad, and how to mitigate the risks.

Non-Technical Skills in the Workplace (20:37) Victoria Silverman is an analyst at Statistics Canada. We discuss the importance of non-technical skills in the workplace and for project work.

Setting-Up as a Consultant (22:14) Oliver Benning is a Masters Student in Computer Engineering at the University of Ottawa, and a consultant with Strikethrough . We discuss how and why he decided to become a quantitative consultant.

Workflow: Predicting Algae Blooms

PROBLEM DESCRIPTION

The ability to monitor and perform early forecasts of various river algae blooms is crucial to control the ecological harm they can cause. The dataset which is used to train the learning model consists of:

- chemical properties of various water samples of European rivers
- the quantity of seven algae in each of the samples, and
- the characteristics of the collection process for each sample.

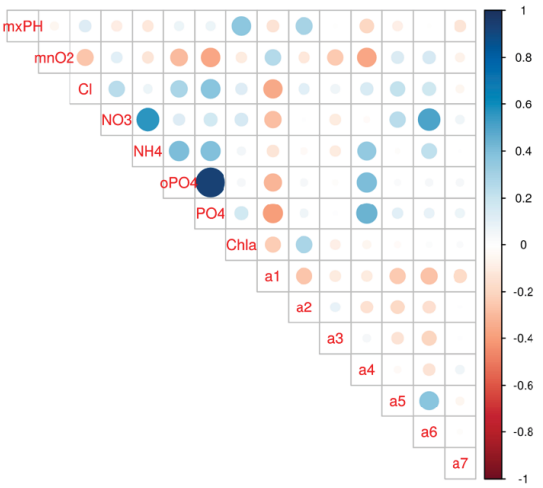
What is the data science motivation for such a model? After all, we can analyze water samples to determine if various harmful algae are present or absent. Chemical monitoring is cheap and easy to automate, whereas biological analysis of samples is expensive and slow. Another answer is that analyzing the samples for harmful content does not provide a better understanding of what drives the production of algae: it just tells us which samples contain algae.

The algae blooms dataset has 338 observations of 18 variables each: *season*, *size*, *speed*, *mxPH*, *mnO2*, *Cl*, *NO3*, *NH4*, *oPO4*, *PO4*, *Chla*, *a1*, *a2*, *a3*, *a4*, *a5*, *a6*, *a7*.

- 3 of the fields are categorical (*season*, *size*, *speed*, which refer to the data collection process)
- of the numerical fields, 8 have names that sound vaguely "chemical"
- the remaining fields refer to various algae blooms

We can get a better feel for the data frame by observing it as an array (first 6 rows):

season	size	speed	mxPH	mnO2	Cl	NO3	NH4	oPO4	PO4	Chla	a1	a2	a3	a4	a5	a6	a7
winter	small	medium	8.00	9.8	60.800	6.238	578.000	105.000	170.000	50.0	0.0	0.0	0.0	0.0	34.2	8.3	0.0
spring	small	medium	8.35	8.0	57.750	1.288	370.000	428.750	558.750	1.3	1.4	7.6	4.8	1.9	6.7	0.0	2.1
autumn	small	medium	8.10	11.4	40.020	5.330	346.667	125.667	187.057	15.6	3.3	53.6	1.9	0.0	0.0	0.0	9.7
spring	small	medium	8.07	4.8	77.364	2.302	98.182	61.182	138.700	1.4	3.1	41.0	18.9	0.0	1.4	0.0	1.4
autumn	small	medium	8.06	9.0	55.350	10.416	233.700	58.222	97.580	10.5	9.2	2.9	7.5	0.0	7.5	4.1	1.0
winter	small	high	8.25	13.1	65.750	9.248	430.000	18.250	56.667	28.4	15.1	14.6	1.4	0.0	22.5	12.6	2.9



A portrait of the relationships between the variables is provided by the correlogram on the left (for the numerical variables).

For now, we assume that the dataset has been properly explored and understood, and that any problems related to invalid data (outliers, etc.) have been solved.

PREDICTION MODELS

Our goal is to build a predictive model for the various algae blooms *a1* – *a7*. It is a supervised learning tasks; in order to mitigate overfitting (a consequence of the bias-variance trade-off), we set aside a test set on which the models (which will be learned on the training set) are evaluated. We use a 65%-35% split (218 – 120 randomly selected training/test observations).

Figure 13.9: Executive summary of the analysis of the algae blooms dataset I.

GENERALIZED LINEAR MODEL

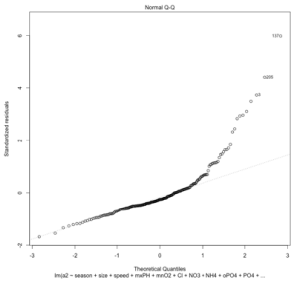
As a baseline model, we run a linear model to predict *a2*, for example, against all the predictor variables, but using only the training set as data. The results are summarized below.

```
Residuals:
    Min      1Q  Median      3Q     Max
-17.436  -5.281  -2.613   2.026  62.712

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -3.083e+01  1.257e+01  -2.452 0.015056 *
seasonsummer -1.166e-01  2.112e+00  -0.055 0.956035
seasonautumn  1.071e+00  2.370e+00   0.452 0.651934
seasonwinter -1.451e+00  2.000e+00  -0.726 0.468935
sizemedium   -2.628e+00  1.895e+00  -1.387 0.166896
sizelarge    -3.210e+00  2.412e+00  -1.331 0.184767
speedmedium   3.887e+00  2.485e+00   1.564 0.119325
speedhigh    -1.104e+00  2.772e+00  -0.398 0.690751
mxPH          4.859e+00  1.559e+00   3.117 0.002092 **
mnO2         -1.841e-01  3.924e-01  -0.469 0.639474
Cl           -7.432e-03  2.006e-02  -0.371 0.711351
NO3           2.132e-01  3.028e-01   0.704 0.482249
NH4          -5.979e-04  5.355e-04  -1.117 0.265510
oPO4          2.290e-03  9.876e-03   0.232 0.816875
PO4          -1.559e-03  5.936e-03  -0.263 0.793090
Chla          1.652e-01  4.614e-02   3.579 0.000432 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.74 on 202 degrees of freedom
Multiple R-squared:  0.206,     Adjusted R-squared:  0.147
F-statistic: 3.493 on 15 and 202 DF,  p-value: 2.498e-05
```

We see that the adjusted R^2 coefficient is fairly small. Furthermore, if the linear model is a good fit, the residuals should have a mean of zero and be "small", which doesn't seem to be the case (at least, relative to the range of *a2*, see 6-pt summary to the right).

```
       a2
Min.   : 0.000
1st Qu.: 0.000
Median : 2.800
Mean   : 7.207
3rd Qu.:10.025
Max.   :72.600
```

The normal QQ-plot for the residuals (see figure on the right), in particular, seem to indicate that linearity of the data is probably not met, as an assumption.



On the other hand, the F−statistic seems to indicate some (linear) dependence on the predictor variables.

Backward elimination stepwise selection suggests that the best linear model for *a2* involves *speed*, *mxPH*, and *Chla*.

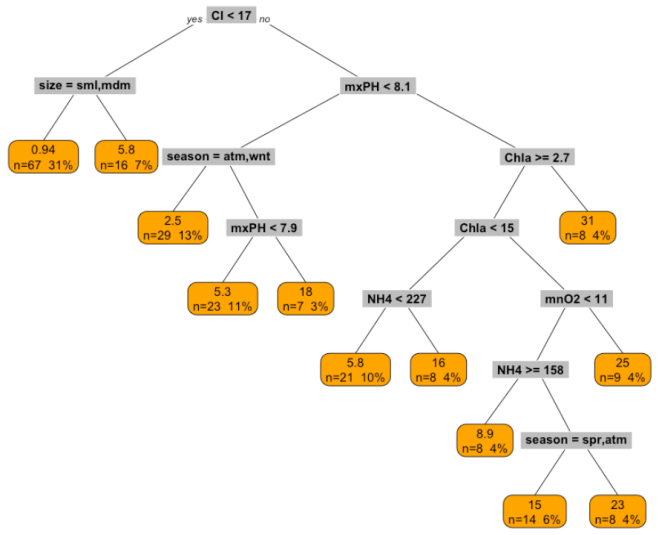
```
Residuals:
    Min      1Q  Median      3Q     Max
-16.195  -6.008  -2.530   2.024  63.589

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -27.13270   11.07921  -2.449 0.015134 *
speedmedium   4.17176    2.34330   1.780 0.076453 .
speedhigh    -0.32929    2.41899  -0.136 0.891850
mxPH          3.89794    1.35358   2.880 0.004387 **
Chla          0.15945    0.04387   3.635 0.000349 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.58 on 213 degrees of freedom
Multiple R-squared:  0.1874,    Adjusted R-squared:  0.1721
F-statistic: 12.28 on 4 and 213 DF,  p-value: 5.289e-09
```

The fit is still not ideal (the value of the adjusted R^2 is quite small).

REGRESSION TREE MODEL

An alternative to regression is the use of regression trees. A recursive partition tree for *a2* is shown below, as is a pruned tree, with the relative importance of the variables for both models:

```
Variable importance
  Chla     NH4      Cl    mxPH    oPO4     PO4     NO3   speed   mnO2 season    size
    19      14      14      13      11       9       6       5       4       3       2
```




```
Variable importance
  Chla     Cl    NH4   mxPH   oPO4    PO4  speed    NO3   mnO2
    19     18     14     13     12     11      7      5      2
```

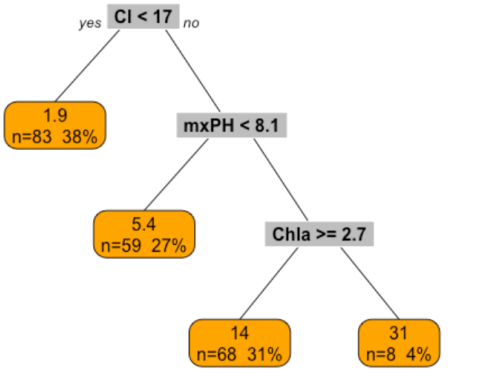


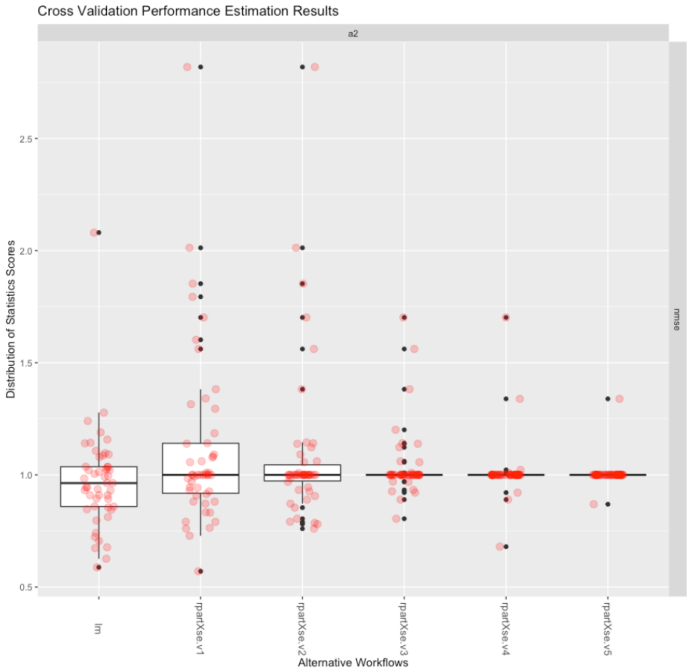
Figure 13.10: Executive summary of the analysis of the algae blooms dataset II.

MODEL EVALUATION

At this stage, we know that the linear model is not great for *a2*, and we have grown regression trees for *a2* but we have not yet discussed whether these models are good fits for *a2*, to say nothing of the remaining 6 algae concentrations.

Various metrics can be used to determine how the predicted values on the test set compare to the actual values: we will use the normalized mean squared error (NMSE). NMSE is unitless: values between 0 and 1 indicate that the model performs better than the baseline; values greater than 1 indicate that the model's performance is sub-par.

The test NMSE for the linear model and for a family of regression tree models (one for 5 different values of a growth/pruning parameter) is estimated using 5 repetitions of 10-fold cross-validation. For each model, the results for the 50 cross-validated models are shown in the image to the right. Summaries for the 50 models for each approach are found below.



lm	nmse	rpartXse.v1	nmse	rpartXse.v2	nmse
avg	0.9880781	avg	1.0333720	avg	1.0596868
std	0.3682616	std	0.3406970	std	0.3147441
med	0.9470239	med	1.0000000	med	1.0000000
iqr	0.2817843	iqr	0.1842643	iqr	0.0435684
min	0.4869917	min	0.6171205	min	0.5049684
max	2.5236216	max	2.4535376	max	2.4535376
rpartXse.v3	nmse	rpartXse.v4	nmse	rpartXse.v5	nmse
avg	1.028517	avg	1.012748	avg	1.001631
std	0.230181	std	0.078035	std	0.011533
med	1.000000	med	1.000000	med	1.000000
iqr	0.000000	iqr	0.000000	iqr	0.000000
min	0.528342	min	0.819828	min	1.000000
max	2.365684	max	1.413850	max	1.081548

It's not necessarily clear which of the models has smaller values of NMSE overall, although it does seem that the latter versions of the regression tree models are not substantially better than the baseline model. The first regression tree model sometimes produces very small NMSE values, but that's offset by some of the larger values it also produces (similarly for the linear model). At any rate, visual evidence seems to suggest that the linear model is the best predictive model for a2 given the training data.

This might seem disheartening at first given how poorly the linear model performed, but it is helpful to remember that there is no guarantee that a decent predictive model even exists. Furthermore, regression trees and linear models are only two of a whole collection of possible models. How do support vector regression or random forests models perform, for instance?

We repeat the task of estimating test NMSE *via* 5 replicates of 10-fold cross-validation for 8 models (linear regression, support vector regression, 3 regression trees, 3 random forests) for all target variables (*a1* – *a7*) simultaneously. We are not looking for a single model which will optimize all learning tasks at once, but rather that we can prepare and evaluate the models for each target variable with the same bit of code. The results are shown in the figure to the right. The top performers (average value of NMSE) for each response are shown on the next page.

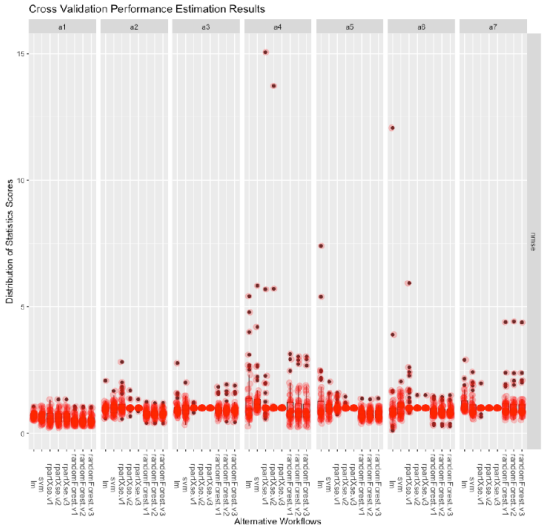


Figure 13.11: Executive summary of the analysis of the algae blooms dataset III.

```
Rank.a1      model  est.nmse
1 randomForest.v2 0.5217204
2 randomForest.v3 0.5228744
3 randomForest.v1 0.5264328
Rank.a5      model  est.nmse
1 randomForest.v1 0.7626241
2 randomForest.v2 0.7675794
3 randomForest.v3 0.7681834
```

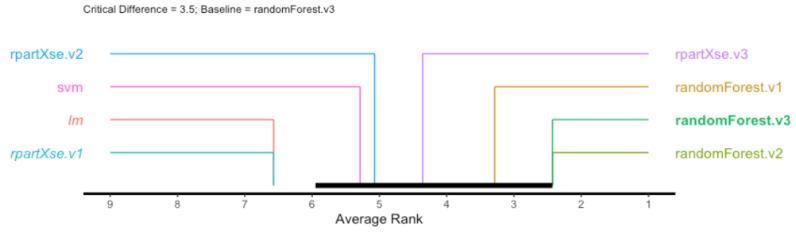
```
Rank.a2      model  est.nmse
1 randomForest.v3 0.7798749
2 randomForest.v2 0.7806831
3 randomForest.v1 0.7849360
Rank.a6      model  est.nmse
1 randomForest.v2 0.8590227
2 randomForest.v3 0.8621478
3 randomForest.v1 0.8663869
```

```
Rank.a3      model  est.nmse
1 randomForest.v3 0.9377108
2 randomForest.v2 0.9400108
3 randomForest.v1 0.9431801
Rank.a7      model  est.nmse
1      rpartXse.v2  1.00000
2      rpartXse.v3  1.00000
3      rpartXse.v1  1.00797
```

```
Rank.a4      model est.nmse
1      rpartXse.v3 1.001453
2 randomForest.v3 1.006496
3 randomForest.v1 1.006806
```

At first glance, the 3rd random forest model (the one that build predictions on 700 trees, as opposed to 200 and 500 for the other random forests models) seems to perform best, but these rankings do not report on the standard error, and so we cannot tell whether the differences between the estimated test NMSEs are statistically significant on the basis of the estimates alone.

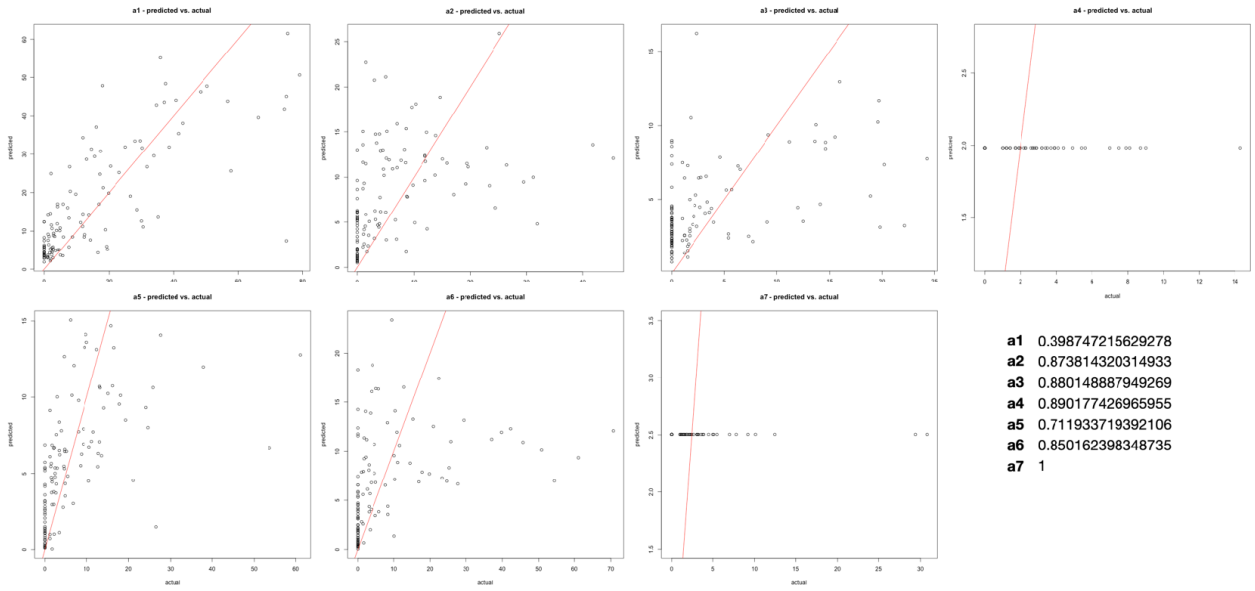
Using the 3rd random forest model as a baseline, we compute the rank differences to the other 7 models for all target variables. The critical rank difference is 3.52. On average, the rank difference to the other models is shown in the list on the right. We can reject with 95% certainty that the performance of the baseline method is the same as that of the linear model and the first regression tree model (`rpartXse.v2`), but not that it is better than the other 5 models. The information is also displayed in the Bonferroni-Dunn CD diagram below.



lm	4.57
svm	2.14
rpartXse.v1	4.28
rpartXse.v2	3.28
rpartXse.v3	2.85
randomForest.v1	1.14
randomForest.v2	0.57

MODEL PREDICTIONS

The best performer for each target response was identified from the cross-validation procedure above: for each target variable *a1* – *a7*, we run the best performer on the original training data to learn a model that is used to predict the appropriate target response for observations in the original test set. Scatterplots of predicted (y-axis) vs. actual levels (x-axis) for test observations are shown below (top: *a1* – *a4*, bottom: *a5* – *a7*), as are the true test NMSEs.



a1	0.398747215629278
a2	0.873814320314933
a3	0.880148887949269
a4	0.890177426965955
a5	0.711933719392106
a6	0.850162398348735
a7	1

Figure 13.12: Executive summary of the analysis of the algae blooms dataset IV.

13.7 Exercises

You may need to consult the videos available on [81] to answer some of the questions. Some of the questions are taken from [92].

1. Now that you know a little bit more about the consulting process, what are two things that appeal to you about it? What are two things that are more off-putting to you?
2. What are some ethical ideals that would guide your choice of quantitative work (employment, contracts, etc.)?
3. Which of the quantitative consulting roles appeals the most to you? The least?
4. Do any of the items on the Consulting Cheat Sheet surprise you? Is there anything that is not on the list that should be added?
5. What does the "Tyranny of Past Success" refer to in the marketing context? How would you try to mitigate against it? Do you think this tyranny also applies in other contexts?
6. How could you address the (potential) differences between your business social media needs and your personal ones?
7. When would you consider sending a prospective client to someone else? What are the advantages of doing so?
8. What small talk would you feel comfortable doing?
9. Among the factors listed in favour of/against working alone or as part of a team, which ones are more compelling to you?
10. How long do you think team meetings should last? How frequently should they occur?
11. What might be some reasons for you to walk away from a project at the contracting stage?
12. What does work-life balance mean to you? What does your Time Management table look like? In re: Hofstadter's Law, what do you think your factor is (how much longer does it take you to do things than you believe it will, on average)?
13. Can you provide an example of a knowledge gap?
14. What would you say your Top 3 quantitative analysis skills are? Where do you need to improve? What would you like to learn more about?
15. Why might the best academic solution not be an acceptable consulting solution?
16. What do you think are the easiest deliverables to prepare? The hardest? The most annoying?
17. Why might some consultants feel that invoicing is problematic?
18. What is the importance of the post-mortem (internal or external)?
19. What do you think are the three most important lessons about clients? Why?
20. What do you think are the three most important lessons about consultants? Why?
21. What is business development? Why should consultants care about business development?
22. What are client's worries in the consulting process? What drives their choice of consultants?
23. What are the components of the trust equation? How important are they to build trust?
24. What trust equation components do you feel that you would need

to improve the most? The least?

25. Write a paragraph explaining why you are taking a quantitative consulting or data analysis course. Were you precise, clear, and accurate? Is this technical writing? Does it need to be?
26. Write a rough outline (with section headers and main ideas) for a blog article on a topic of your choice. Keep in mind that the document's organization is dependent on the target audience.
27. Revise the vague words and phrases, specifying exact information. Invent numbers and modify the rest of the sentences as needed.

    a. I have a *low GPA*.
    b. The b-ball player was *really tall*.
    c. I'll be home *as soon as possible*.
    d. The team has a *losing record*.
    e. The computer has *lots of memory*.

28. Change the following long words to shorter words.

    a. utilize
    b. anticipate
    c. cooperate
    d. indicate
    e. initially
    f. presently
    g. prohibit
    h. inconvenience

29. Change the following phrases to one word.

    a. in the event that
    b. at this point in time
    c. with regard to
    d. in the first place
    e. is of the opinion that
    f. due to the fact that
    g. make revisions
    h. take into consideration
    i. with the exception of
    j. make an adjustment to/of

30. Revise the following long sentences, making them shorter.

    a. I will be calling you on May 31 to see if you have any questions at that time.
    b. If I can be of any assistance to you in the evaluation of this proposal, please feel free to give me a call.
    c. The company is in the process of trying to cut the cost of expenditures relating to the waste of unused office supplies.
    d. I am of the opinion that graduate students have too much work to do.
    e. In the month of July, my family will make a visit to the province of New Brunswick.
    f. It is the company's plan to take action to avoid problems with hazardous waste.
    g. On two separate occasions, the manager of personnel met with at least several different employees to ascertain whether or not they were in agreement with the company's policies regarding overtime.

31. Reformat the following text by using highlighting techniques. Consider using bullets or numbers, headings, boldface or underlining, and white space.

    > To make a pie chart using your word processing package's graphic components, turn on the machine. Once it has booted up, double click on the word processing icon. After the system is open, click on "graphic," scroll down to "chart," and double click. Next, click on "data chart types" and select "pie." Once you have done this, input your new data in the "data sheet." After this has been completed, click anywhere on the page to import your new pie chart. If you want to make changes, just double click again inside the pie chart; then you can revise according to your desires.

32. Make a list of 4-6 acronyms or abbreviations from an area of interest. What percentage of the audience understand your acronyms? Define / explain the terms for low-tech peers and for lay readers.
33. Describe your footwear, as accurately as possible. Without knowing the purpose of the task, how difficult is it to know how long or how specific you should be?
34. Offer a technical writing critique of this chapter (there are numerous inconsistencies).