Technical writing is different from expository or literary writing in a number of ways. I will briefly discuss

Structure
Figures, tables, and equations
Writing about data
Writing about statistical calculations

More stuff to read (chapters on github repo):
- Appendix to “Stat Labs” by Deb Nolan and Terry Speed
- Book “Eloquent Science” by David M. Schultz - see especially the chapter for non-native speakers of English
- Article “The Science of Scientific Writing” by Gopen and Swan - introduces idea of reader expectations

Examples for this lecture are taken from

Beaudoin and Swartz (2010) Strategies for pulling the goalie in hockey. The American Statistician, 64: 197-204.

Berrocal et al. (2010) Probabilistic weather forecasting for winter road maintenance. Journal of the American Statistical Association, 105: 522-537.

McShane and Wyner (2011) A statistical analysis of multiple temperature proxies: Are reconstructions of surface temperatures over the last 1000 years reliable? The Annals of Applied Statistics, 5: 5-44.

Wolkewitz et al. (2010) Two pitfalls in survival analyses of time-dependent exposure: A case study in a cohort of Oscar nominees. The American Statistician, 64: 205-211.

Typical sections in a statistics report or paper:

Title
Author(s)
Abstract / executive summary
Keywords
Introduction
Background / literature review
Data description
Methods
Results
Discussion / conclusion
Appendices
References

However, this varies *widely*....

Examples:

Beaudoin & Swartz

Title
Authors
Abstract
Key words
1. Introduction
2. Data analysis
3. Simulation model
4. Bayesian parameter estimation
5. Simulation results
6. Concluding remarks
References

Berrocal et al.

Title
Authors
Abstract
Key words
1. Introduction
2. Data and methods
    2.1 Road maintenance problem
    2.2 Statistical model
    2.3 Model fitting
    2.4 Choice of training period
    2.5 Generating forecasts
3. Results
4. Discussion
References

McShane & Wyner

Some observations:

The number of sections and levels in the hierarchy is related to the length of the paper; readers need more "signposts" in a longer paper. But it's rare to have a nested structure with more than two levels.

The material often determines its own logical structure, although some topics must come before others (e.g. methods before results).

It helps to highlight difficult and/or important topics in their own section or sub-section. Informative names make it easy for readers to find what they need.

Important: It is very rare that technical reports or journal articles are read through from start to finish, in order. Most readers employ non-linear reading and skimming.

A more typical ordering for reading would be

Title, Abstract
Skim Figures, Introduction, Results, Conclusions
Skim Methods
Perhaps decide to read in detail

A study of managers at Westinghouse showed 100% read the abstract, 60% read the introduction, 50% read the conclusion, 15% read the body of the report, and 10% read the appendix.

You can use non-linear reading to your advantage by

1. Writing the abstract carefully to reflect the main findings and highlight what is interesting.

2. Choosing your figures to reflect your main points, making sure they are clear, and including important conclusions drawn from the figure in a caption as well as in the text.

3. Anticipating what the reader will be looking for, and making these things into specific section headings.

You might also consider writing in a non-linear order.

For example, my typical order of writing a paper is something like

- Figures & tables
- (Outlining)
- Methods & results
- Intro & discussion
- (More analysis)
- Revisions to above
- Abstract

It really helps to start writing BEFORE your analysis is completely done. This helps highlight exactly what is still needed.

Figures and tables should be numbered and referred to by number in the text. The caption should include any details needed to interpret the plot.

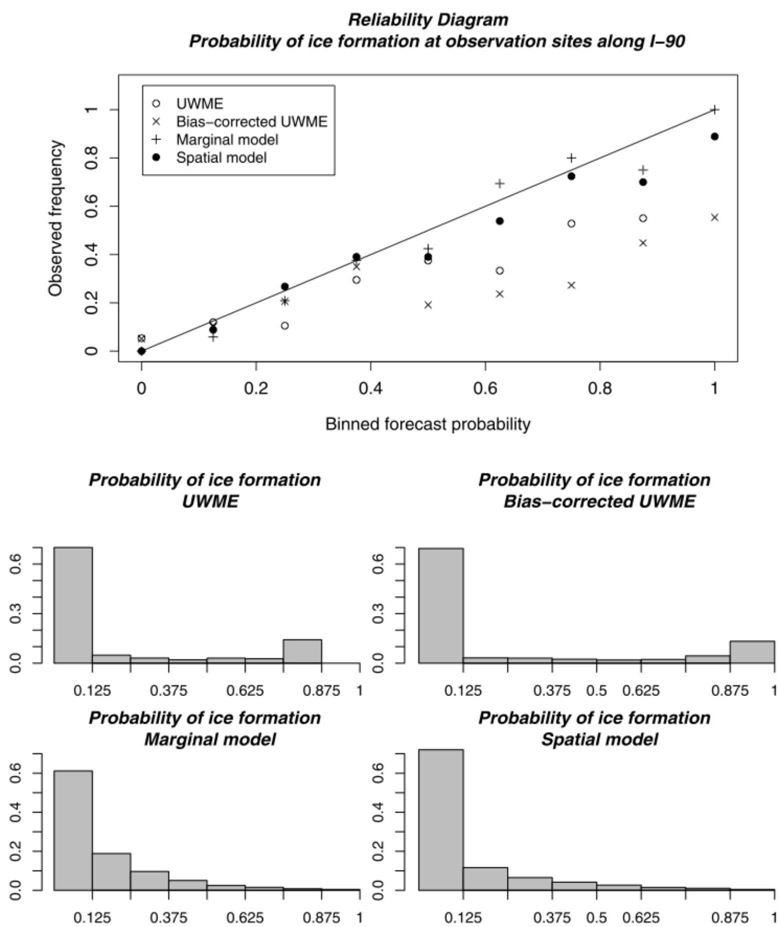

Figure 7. Reliability diagram for probability forecasts of ice formation at observation sites by the UWME, the bias-corrected UWME, the marginal model, and the spatial model, for the 2003–2004 and 2004–2005 winter seasons. Histograms of the forecast probabilities are also shown.

Table 2 shows that for the probability that ice forms at individual locations, both of our probabilistic forecasting methods substantially outperformed the raw and the ensemble forecasts. Figures 7 and 8 show reliability diagrams for the probabilistic forecasts of ice formation provided by the marginal model, the spatial model, the UWME, and the bias-corrected UWME. Our methods were superior to both ensembles, which were very underdispersed and particularly unreliable for high predicted probabilities. At individual locations along I-90, the marginal and spatial methods performed similarly, as expected. However, for probability forecasts of the spatial aggregate "ice formation along $\mathcal{I}$," the spatial model was more reliable: the marginal model tended to overestimate the probability of ice formation.

Note that plots in journal articles or reports often write in the caption how different lines or symbols are to be interpreted. This convention is not universal (see e.g. the figure on the previous slide, which used a legend).

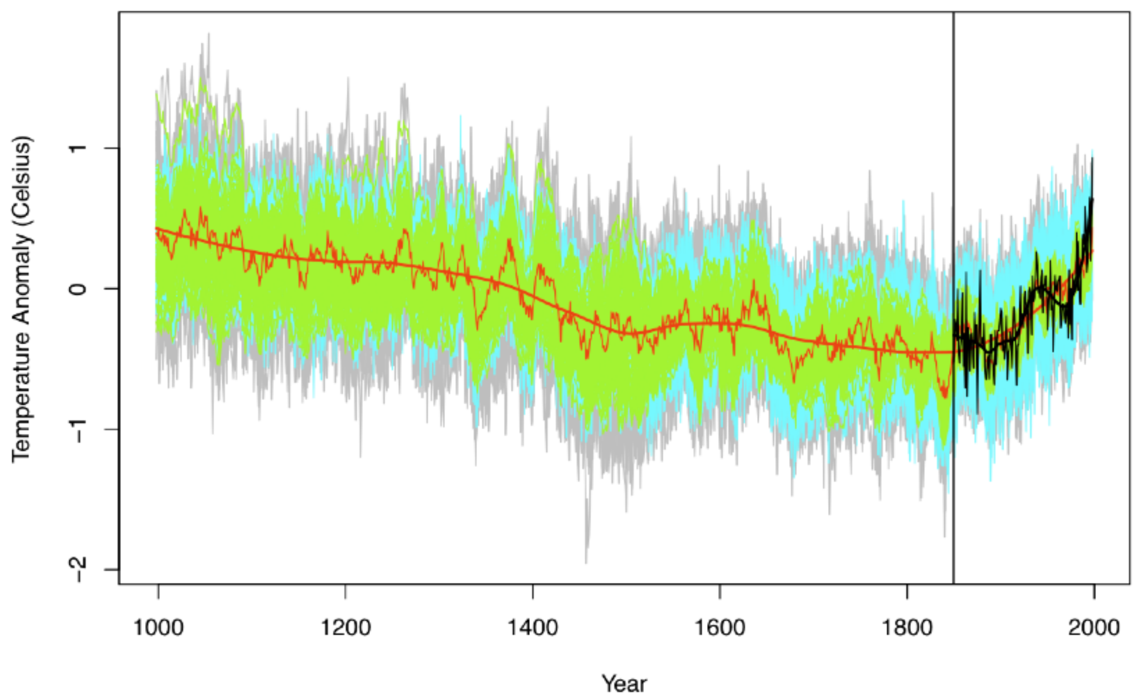

FIG 16. *Backcast from Bayesian Model of Section 5. CRU Northern Hemisphere annual mean land temperature is given by the thin black line and a smoothed version is given by the thick black line. The forecast is given by the thin red line and a smoothed version is given by the thick red line. The model is fit on 1850-1998 AD and backcasts 998-1849 AD. The cyan region indicates uncertainty due to* ϵ_t*, the green region indicates uncertainty due to* $\vec{\beta}$*, and the gray region indicates total uncertainty.*

Mathematical expressions are sometimes separated from the rest of the text for emphasis, particularly if they are long. They are numbered only if they are referred to at some later point in the text.

Our model assumes that individual team scoring rates arise from a population of league-wide scoring rates

$$\theta_{is} \sim \text{Gamma}(a_s, b_s),$$

where the parameters a_s and b_s have independent prior distributions

$$a_s \sim \text{Gamma}(\alpha_{as}, \beta_{as}) \quad \text{and} \\ b_s \sim \text{Gamma}(\alpha_{bs}, \beta_{bs}). \tag{3}$$

The hyperparameters α_{as}, β_{as}, α_{bs}, and β_{bs}, $s = 1, \ldots, \tilde{m}$, are set in an empirical Bayes fashion by considering the sample scoring rates. The Gamma hyperparameters are chosen such that $\alpha_{as} > 1$ and $\alpha_{bs} > 1$. We impose a Uniform(0, 1) prior for f in (2) according to the widely held belief that home-ice confers an advantage. The primary parameter of interest in our

When **presenting the data**, ask yourself:

Who collected the data and is it publicly available? Often a URL is included. If the data were collected as part of an experiment, describe how this was conducted.

What are the variables (only those included in the report) and their units? Summary statistics and plots may be used to highlight important features.

Is there missing data? If so, how much, and how will this be treated?

Is there a particular date range or location(s) associated with the data?

When **writing about calculations**, you should describe them (giving mathematical expressions where appropriate), but generally do NOT include code or names of specific variables or functions in the report.

Examples:

"We calculated MLEs for each of B = 1000 simulated datasets according to (5). Histograms of their distributions are shown in Figure 3."

"All calculations were carried out in the R programming language, with additional functions from the fields package. Annotated code is contained in the Appendix."