INTRODUCTION TO DATA ANALYSIS

Contents

Preface 5

0.1 Testing / Showcasing 5

1 *General Introduction* 7

2 *Data* 9

3 *Probability* 11

4 *Models* 13

5 *Inference* 15

6 *Hypothesis Testing* 17

7 *Model Comparison* 19

8 *Generalized Regression Modeling* 21

Preface

The book introduces key concepts of data analysis from a frequentist and a Bayesian tradition. It uses R to handle, plot and analyze data. It relies on simulation to illustrate selected statistical concepts.

0.1 Testing / Showcasing

Don't pay too much attention to what is written here.

0.1.1 Quotes

This is a quote:

> Tidy datasets [...] have a specific structure: each variable is a column, each observation is a row, and each type of observational unit is a table.
>
> — Wickham (2014)

0.1.2 Infobox

At certain stages, possibly at the end of chapters or after important concepts, we might want to use a special infobox (see `.infobox` in `styles.css`) to summarise it or give food for thought. Like this:

A horse walks into a bar and orders a pint. The barkeep says "you're in here pretty often. Think you might be an alcoholic?", to which the horse says "I don't think I am.", and vanishes from existence.

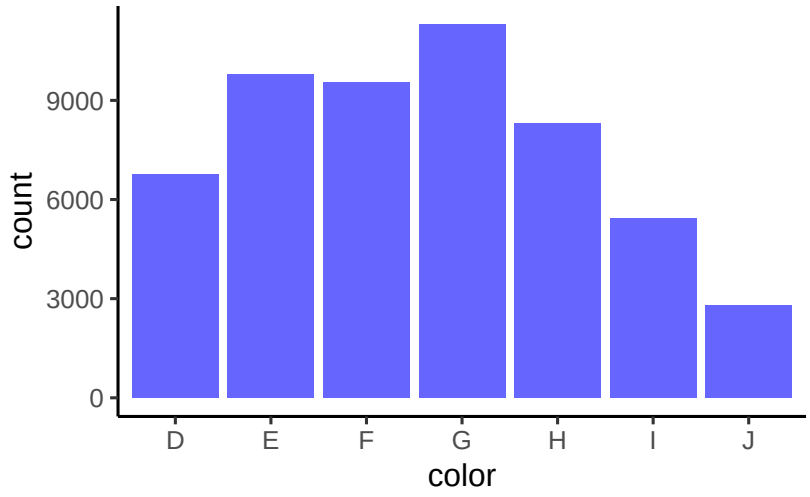
See, the joke is about Descartes' famous philosophy of 'I think therefore, I am", but to explain that part before the rest of the joke would be to put Descartes before the horse.

0.1.3 Plots

This is a plot:

```
library(ggplot2)

ggplot(diamonds, aes(color)) +
  geom_bar(fill = "blue", alpha = .6) +
  theme_classic()
```



1
General Introduction

- what stats is about
- different practices
- learning goals

2
Data

learning goal: how to arrange, summarize and visualize (aspects of data) to address a question of interest ("hypothesis-driven data poking")

- different kinds of data
- summary statistics
- data wrangling
- data plotting

3
Probability

learning goal: get comfortable with basic notions of probability theory

- probability distributions
- random variables
- conditional probability
- selected distributions

4
Models

learning goal: diagnosing the (conceptual) differences between kinds of statistical models

- priors & likelihood
- conceptual differences between frequentist and Bayesian approaches (revisited)
- notation (probabilistic causal networks)
- three example models:
  - "binomial model"
  - "factorial-design model"
  - simple linear regression model

5
Inference

- MLE vs posterior
- confidence intervals
- credible intervals
- briefly: algorithms for MLE & Bayesian inference

6
Hypothesis Testing

- binomial test
- t-test
- ANOVA
- linear regression

7
Model Comparison

- AIC
- likelihood ratio test
- Bayes factor

8
Generalized Regression Modeling

- applications

Wickham, Hadley. 2014. "Tidy Data." *Journal of Statistical Software, Articles* 59 (10): 1–23. https://doi.org/10.18637/jss.v059.i10.