Introduction to Data Analysis

last rendered at: 2019-09-06 09:37:02

Contents

Preface **5**
0.1 Testing 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

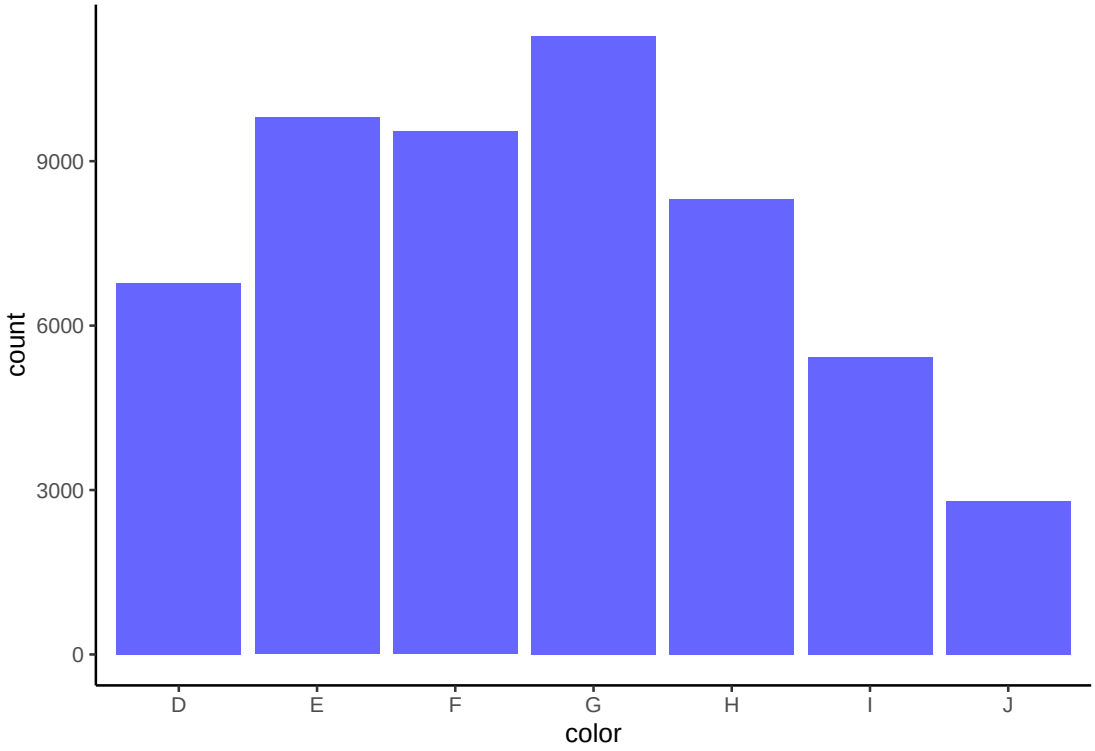
This is a quote by Wickham (2014):

> 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.

This is a plot:

```
library(ggplot2)

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


9000
6000
3000
0
count
D
E
F
G
H
I
J
color

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.