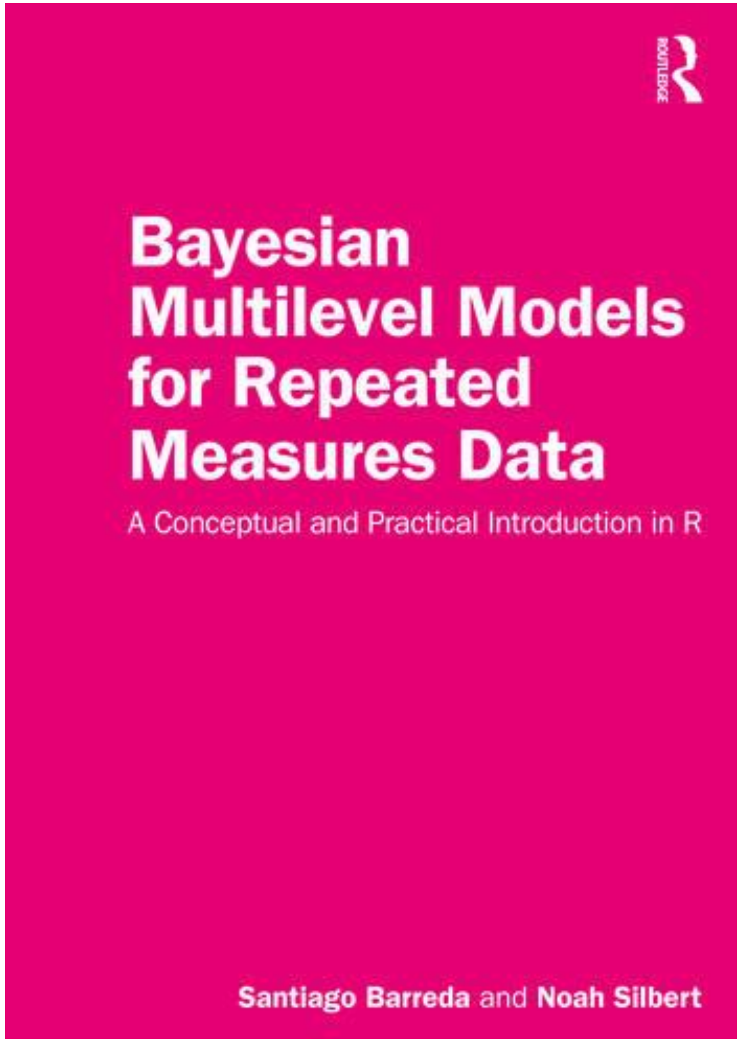
Preface & Chapter 1

Lin 205C
Santiago Barreda

Course Textbook



Course Outline

Week	Date	Topics
1	Sep. 27	Chapter 1: Introduction: Experiments and Variables
2	Oct. 4	Chapter 2: Probabilities, likelihood, and inference
3	Oct. 11	Chapter 3: Fitting Bayesian regression models with brms
4	Oct. 18	Chapter 4: Inspecting a 'single group' of observations using a Bayesian multilevel model
5	Oct. 25	Chapter 5: Comparing two groups of observations: Factors and contrasts
6	Nov. 1	Chapter 6: Variation in parameters ('random effects') and model comparison
7	Nov. 8	Chapter 7: Comparing many groups, interactions, and posterior predictive checks
8	Nov. 15	Chapter 9: Quantitative predictors and their interactions with factors
9	Nov. 22	Chapter 10: Logistic regression and signal detection theory models
10	Nov. 29	Thanksgiving
11	Dec. 6	Chapter 13: Writing up Experiments

Motivation

- Leave out common information: No boring ‘origin story’.
- All repeated measures, all the time.
- Only useful examples, and present complex examples.
- No frequentism, only Bayesian.

Other Books

Preface ix

1 Preliminaries 1
1.1 Our R Toolset 1
1.2 Our Approach 1
1.3 Context 3

2 Samples, estimates, and hypothesis tests 7
2.1 Preliminaries 7
2.2 Point Estimation 9
2.3 Uncertainty and Interval Estimation 14
2.4 Hypothesis Testing 19
2.5 Parametric and Nonparametric Tests 27
2.6 Common Misconceptions about *p*-Values 33
2.7 Reporting Hypothesis Tests 34
2.8 Other Reading 36
Exercises 37

3 Effect Size, Power, and Error 39
3.1 Preliminaries 39
3.2 Effect Size 40
3.3 Type I/II Errors and Power 47
3.4 Error in Effect Size: Type M and Type S Errors 59
3.5 Assumptions of Hypothesis Tests and Consequences 63
3.6 Other Reading 68
Exercises 68

4 Linear regression 1 71
4.1 Preliminaries 71
4.2 Regression: General introduction 72
4.3 Simple Linear Regression 75
4.4 Multiple Linear Regression 83
4.5 Interactions 87
4.6 Reporting a Linear Regression Model 93
4.7 Other Reading 95
Exercises 95

5 Linear Regression 2 97
5.1 Preliminaries 97
5.2 Linear Regression Assumptions 98
5.3 Problems with the Errors 100
5.4 Problems with the Model 104
5.5 Transformations 108
5.6 Problems with Predictors 114
5.7 Problems with Observations 123
5.8 Trade-offs between Models 128
5.9 Model Comparison 132
5.10 Variable Selection 138
5.11 Other Reading 147
Exercises 148

6 Categorical Data Analysis and Logistic Regression 149
6.1 Preliminaries 149
6.2 Categorical Data Analysis 151
6.3 Odds and Odds Ratios 156
6.4 Simple Logistic Regression 160
6.5 Inference for Logistic Regression 165
6.6 Goodness of Fit 168
6.7 Multiple Logistic Regression 170
6.8 Model Validation 178
6.9 Reporting and Summarizing 185
6.10 Other Readings 189
Exercises 190

7 Practical Regression Topics 191
7.1 Preliminaries 191
7.2 Multilevel Factors: Contrast Coding 193
7.3 Omnibus and Post hoc Tests 212
7.4 Interpreting Interactions 218
7.5 Nonlinear Effects 227
7.6 Collinearity Diagnostics Revisited 238
7.7 Other Readings 239
Exercises 239

8 Mixed-Effects Models 1: Linear Regression 241
8.1 Preliminaries 241
8.2 Motivation: Grouped Data 242
8.3 Linear Mixed Models: Introduction 244
8.4 Random Slopes 256
8.5 Hypothesis Tests 267
8.6 Model Summaries 278
8.7 Random-Effect Correlations 280
8.8 Model Predictions 291
8.9 Reporting the Fitted Model 297
8.10 Model Validation 298
8.11 Other Readings 309
Exercises 310

9 Mixed-Effects Models 2: Logistic Regression 313
9.1 Preliminaries 313
9.2 Introduction 315
9.3 Two Grouping Factors, Random Intercepts and Slopes 321
9.4 Hypothesis Tests 327
9.5 Model Summaries 332
9.6 Model Validation 337
9.7 Nonlinear and Factor Effects 340
9.8 Variable Importance 350
9.9 Reporting a Mixed-Effects Logistic Regression 354
9.10 Other Readings 354
Exercises 355

10 Mixed-Effects Models 3: Practical and Advanced Topics 357
10.1 Preliminaries 357
10.2 More on Random Effects 360
10.3 Model Convergence 364
10.4 Singular Models 374
10.5 Model Selection 379
10.6 Predictions and Uncertainty for Individual Levels 395
10.7 Nonlinear Effects 403
10.8 Power for Mixed-Effects Models 405
10.9 Other Readings 407
Exercises 407

A Appendix: Datasets 409
A.1 `transitions` 409
A.2 `vot` 409

B Appendix: R packages 411

Selected Abbreviations 413
Topic Index 415
Function Index 425
Bibliography 429

Motivation

- Learn by doing.

- Practice

- <u>Statistics is procedural not declarative knowledge</u>.

Why Bayesian Modeling?

- It works whenever frequentist approaches work, and many times when they don’t.

- You get a lot more information from Bayesian models.

- Frequentist approaches offer off-the shelf, one-size-fits all hypothesis testing machines.

- Bayesian modelling let’s you build the model you want.

Class Structure

- One chapter per week.
- Quiz on each chapter.
- Short assignment due before next class.
- Final paper.

Chapter 1

Introduction: Experiments and Variables

Lin 205C
Santiago Barreda

Experiments and Science

- Experiments: Procedures to help answer some research question.

- Experiments are scientific when they adhere to the ‘scientific method’*.

‘The’ Scientific Method

1. Ask questions based on gaps in their knowledge about the world.

2. Collect data using codified procedures developed to avoid certain pitfalls and maximize the chance that the collected data can answer their questions.

3. Evaluate their questions in light of their data.

4. Reach conclusions, where possible, and synthesize their conclusions with their previous knowledge about the world.

Modern Science = Math

- Mostly empirical, and results in the collection of a lot of numbers.
- We then need to describe and quantify patterns in those numbers.

"[The universe] cannot be read until we have learnt the language and become familiar with the characters in which it is written. It is written in mathematical language, and the letters are triangles, circles and other geometrical figures, without which means it is humanly impossible to comprehend a single word." - Galileo

"When you can measure what you are speaking about, and express it in numbers, you know something about it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely, in your thoughts advanced to the stage of science."- Lord Kelvin

Modern Science = ~~Math~~ Statistics

“As far as the laws of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality.” - Einstein

- We can very rarely be exact and deterministic in our knowledge gathering.
- Statistics allows us to think about the amount of uncertainty or variability in the conclusions we reach.

Controlled Experiments and Effects

- A controlled experiment is one where the experimenter interferes to ensure the fairness of an experiment.
  - Think of ‘control’ as continuous rather than discrete.

- Effects: Associations between changes in experimental conditions and outcomes.

- Effects may be causal but may not be.

Controlled Experiment: Example

- Two groups of listeners are asked to read a passage.
- One is given caffeine and the other is not. We check for a difference in average reading times.
- If there is, we conclude caffeine has ‘an effect’ on reading times.
- Does the experiment ‘prove’ this is true? Can anything?

Experiments and Inference

- Inference: Going from premises to conclusions.

- Deduction: Arguments whose conclusions must be true if the premises are true.

- Induction: Arguments whose conclusions may be true if the premises are true.

Induction

- Induction = Probabilistic reasoning

- Most reasoning is inductive

- Almost all experiments are indcutvie.

Problems with Induction I

- The problem of induction: Induction works if and only if the things we did not observe are just like those we did, i.e. if the future is like the past.

“Domestic animals expect food when they see the person who usually feeds them. We know that all these rather crude expectations of uniformity are liable to be misleading. The man who has fed the chicken every day throughout its life at last wrings its neck instead, showing that more refined views as to the uniformity of nature would have been useful to the chicken.” – Bertrand Russel

Problems with Induction II

- Affirming the consequent: Taking a true if/then statement and flipping it.

Example 1: If you live in Davis, then you Live in California.
[wrong] Joe lives in California, therefore, Joe lives in Davis.

Example 2: If caffeine speeds up reading times, the caffeine group will read faster.
[wrong] The caffeine group read faster, therefore, caffeine speeds up reading times.

Example 3: If some linguistic theory is true, I will observe some result.
[wrong] I observed some result, therefore, some linguistic theory is true.

Problems with Induction III

- Much of modern science may consist of affirming the consequent (especially in linguistics!).

- It is difficult/impossible to ‘prove’ a general truth based on limited observations.

- Let’s be humble!

The problem of 'inverse probability'

Probability of data given some hypothesis

$$P(D|H)$$

$$P(H|D)$$

Probability of hypothesis given some data

$$P(H|D) \neq P(D|H)$$

Basic Probability

$P(A)$ = Probability that A is true

$P(B)$ = Probability that B is true

$$P(A \,\&\, B) = P(B|A) \cdot P(A)$$

$$P(B \,\&\, A) = P(A|B) \cdot P(B)$$

How likely is a very tall man to play in the NBA?

- Here is some relevant information:
  - 100,994,367 males over 18 in the USA
  - 3,199 men over 6’ 10” in USA
  - 486 Active NBA players
  - 88 players in the NBA are over 6’ 10”

How likely is a very tall man to play in the NBA?

Marginal ('overall') probabilities

$P(Tall)$ = Probability that a man is over 6'10"

$P(NBA)$ = Probability of playing in the NBA

Joint probabilities

$P(Tall\ \&\ NBA)$ = Probability of being tall AND playing in the NBA

$P(NBA\ \&Tall)$ = Probability of playing in the NBA AND being tall.

Conditional ('if') probabilities

$P(Tall|NBA)$ = Probability of being tall <u>given/if</u> that you play in the NBA

$P(NBA|Tall)$ = Probability of playing in the NBA <u>given/if</u> that you are tall.

How likely is a very tall man to play in the NBA?

of Tall adult males / # of adult males in the USA

$P(Tall) = 3199 / 100{,}994{,}367 = 0.000032$

# of NBA Players / # of adult males in the USA

$P(NBA) = 486 / 100{,}994{,}367 = 0.0000048$

# of Tall, adult male NBA players / # of Tall adult males

$P(NBA|Tall) = 88 / 3199 = 0.028$

# of Tall, adult male NBA players / # of NBA players

$P(Tall|NBA) = 88 / 486 = 0.18$

Conditional Probabilities = Not Reversible

<u>NOT</u> reversible!

$P(NBA|Tall)$ = 88 / 3199 = 0.028 $\neq$ $P(Tall|NBA)$ = 88 / 486 = 0.18

<u>UNLESS</u> you consider the base rate!

$P(NBA|T) \cdot P(Tall) = 0.028 \cdot 0.000032 = 0.00000087$

$P(Tall|NBA) \cdot P(NBA) = 0.18 \cdot 0.0000048 = 0.00000087$

Why??

- This inequality is basic logic: $P(NBA|Tall) \neq P(Tall|NBA)$.

- Why does the 'base rate' fix this?

$$P(NBA|Tall) \cdot P(Tall) = P(Tall|NBA) \cdot P(NBA)$$

$$P(NBA\ \&\ Tall) = P(Tall\ \&\ NBA)$$

Is this a Lion?

Data: The visual image.

Hypothesis: This is not a dog, it’s a Lion.

Is this a Lion?

$$P(\text{Image}|\text{Lion})$$

$$P(\text{Lion}|\text{Image})$$

$$P(\text{Image}|\text{Lion}) \neq P(\text{Lion}|\text{Image})$$

Is this a Lion?

$$P(\text{Lion}|\text{Image}) \neq P(\text{Image}|\text{Lion})$$

$$P(\text{Lion}|\text{Image}) \cdot P(\text{Image}) = P(\text{Image}|\text{Lion}) \cdot P(\text{Lion})$$

$$P(\text{Lion}|\text{Image}) \cancel{\cdot P(\text{Image})} \propto P(\text{Image}|\text{Lion}) \cdot P(\text{Lion})$$

The Bayesian Approach

- Let's say that $P(\text{Image}|\text{Lion})$= 0.8.

- Does this mean you will think this is a lion?

- Does this mean that $P(\text{Lion}|\text{Image})$ is high?

- What should P(Lion) be?

The Bayesian Approach

- On campus, P(Lion) = 0.0001.
- On an African safari, P(Lion) = 0.1.
- So:

$$P(\text{Lion}|\text{Image}) \cancel{\cdot P(\text{Image})} \propto P(\text{Image}|\text{Lion}) \cdot P(\text{Lion})$$

- Is:

- $P(\text{Lion}|\text{Image}) \propto 0.8 \cdot 0.0001 \propto$ 0.0008 on campus.
- $P(\text{Lion}|\text{Image}) \propto 0.8 \cdot 0.1 \propto$ 0.08 on safari.

The Bayesian Approach

- The same evidence leads to different conclusions in different situations.

- This is how we reason “in real life”!

- Bayesian reasoning helps us work “backwards” from data to hypothesis!

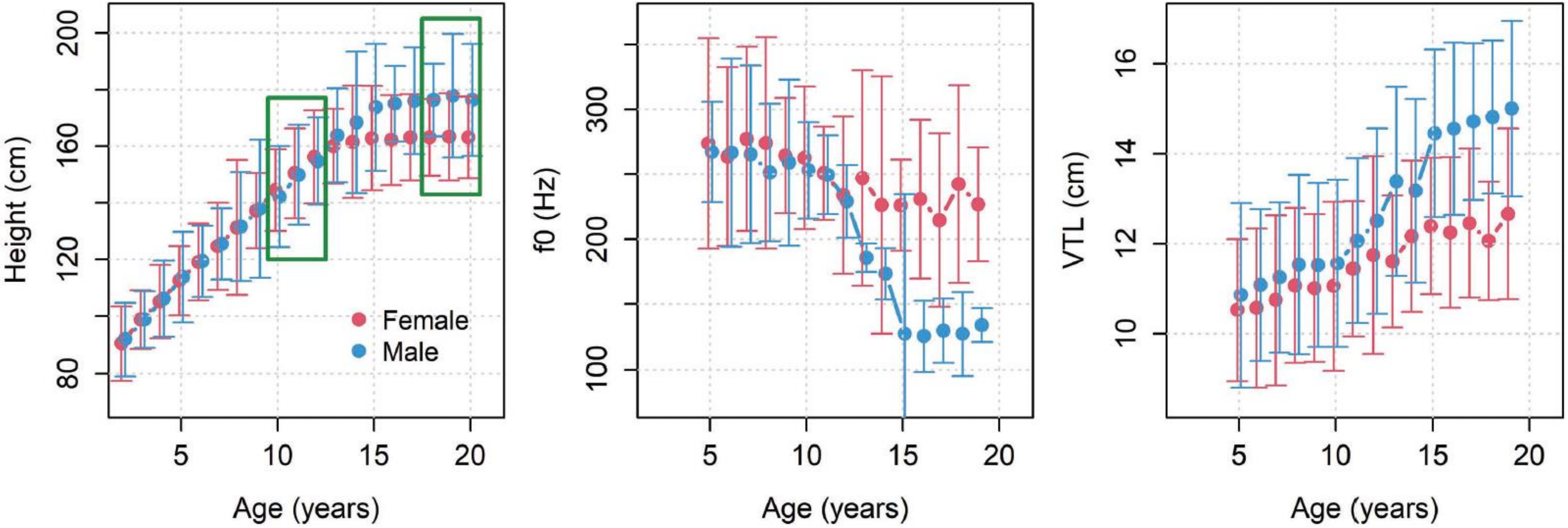
Our Experiment

- An investigation of <u>apparent</u> height, age, and Gender.

- Examining the relationship between apparent speaker characteristics and speech acoustics.

- We played people speech stimuli and collected behavioral measures.

Our Experiment



Our Experimental Methods

- Listeners were played recordings of 139 boys, girls, men, and women saying the word ‘heed’.

- Were asked to judge:

  - The height of the speaker in feet and inches.

  - The gender of the speaker (male or female).

  - The age of the speaker (child 10-12, or adult >18).

Our Experimental Data

```
# load package
library ("bmmb")

# load data
data (exp_data_all)
```

```
# see first 6 rows
head (exp_data_all)
##    L C height R S C_v  vtl  f0 dur G A G_v A_v
## 1 1 g  165.6 a 1   b 12.2 277 237 f c   m   c
## 2 1 w  173.2 b 1   b 12.2 277 237 f a   m   c
## 3 1 w  165.6 a 2   b 12.4 287 317 f a   m   c
## 4 1 g  147.8 b 2   b 12.4 287 317 f c   m   c
## 5 1 g  165.6 a 3   b 11.6 219 277 f c   m   c
## 6 1 g  158.8 b 3   b 11.6 219 277 f c   m   c
```

Our Experimental Data

```
str (exp_data_all)
## 'data.frame':    4170 obs. of  13 variables:
##  $ L     : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ C     : chr  "g" "w" "w" "g" ...
##  $ height: num  166 173 166 148 166 ...
##  $ R     : Factor w/ 2 levels "a","b": 1 2 1 2 1 2 1 2 1 2 ...
##  $ S     : int  1 1 2 2 3 3 4 4 5 5 ...
##  $ C_v   : Factor w/ 4 levels "b","g","m","w": 1 1 1 1 1 1 1 1 1 1 ...
##  $ vtl   : num  12.2 12.2 12.4 12.4 11.6 11.6 11.9 11.9 12.1 12.1 ...
##  $ f0    : int  277 277 287 287 219 219 260 260 244 244 ...
##  $ dur   : int  237 237 317 317 277 277 318 318 242 242 ...
##  $ G     : chr  "f" "f" "f" "f" ...
##  $ A     : chr  "c" "a" "a" "c" ...
##  $ G_v   : Factor w/ 2 levels "f","m": 2 2 2 2 2 2 2 2 2 2 ...
##  $ A_v   : Factor w/ 2 levels "a","c": 2 2 2 2 2 2 2 2 2 2 ...
```

Our Experimental Data

```r
# show the first six
head(exp_data_all$height)
## [1] 165.6 173.2 165.6 147.8 165.6 158.8

# show the first element
exp_data_all$height[1]
## [1] 165.6

# show elements 2 to 6
exp_data_all$height[2:6]
## [1] 173.2 165.6 147.8 165.6 158.8
```

```r
head(exp_data_all[[3]])
## [1] 165.6 173.2 165.6 147.8 165.6 158.8

head(exp_data_all[["height"]])
## [1] 165.6 173.2 165.6 147.8 165.6 158.8
```

Our Experimental Data

```
head(exp_data_all[,3])
## [1] 165.6 173.2 165.6 147.8 165.6 158.8

head(exp_data_all[,"height"])
## [1] 165.6 173.2 165.6 147.8 165.6 158.8
```

```
exp_data_all[1,]
##   L C height R S C_v  vtl  f0 dur G A G_v A_v
## 1 1 g  165.6 a 1   b 12.2 277 237 f c   m   c

exp_data_all[1,2]
## [1] "g"
```

Variables

- Placeholders for some value (known or unknown).

- Random variables: Variables whose values aren’t known a priori (before observation).

Populations and Samples

- The population: The entire set of values or possible outcomes.

- The sample: The set of outcomes/values you observe.

Dependent and Independent Variables

- Dependent/outcome variables: Variables you want to understand.

- Independent/predictor variables: Variables you use to explain.

Types of Variables

- Quantitative: Numeric values on at least an internal scale. Usually a large number of possible outcomes.

- Categorical/factors: Non-numeric values, usually a ‘smaller’ number of possible outcomes. Possible values are called levels.

- Ordinal: Categorical variables the preserve an order but not a distance between elements.

Categorical Variables

```
# see the first 6 observations
head (exp_data_all$C_v)
## [1] b b b b b b
## Levels: b g m w

# it has levels
levels(exp_data_all$C_v)
## [1] "b" "g" "m" "w"

# each level has numerical values
table (exp_data_all$C_v, as.numeric (exp_data_all$C_v))
##
##          1    2    3    4
##    b   810    0    0    0
##    g     0  570    0    0
##    m     0    0 1350    0
##    w     0    0    0 1440
```

Continuous Variables?

- Is the variable on a ratio or interval scale? This is a prerequisite for a quantitative value to be used as a dependent variable. An interval scale means that differences between values are meaningful, and a ratio scale additionally means that 0 is meaningful.
- Is the underlying value continuous? Many variables are discrete in practice due to limitations in measurement. However, if the underlying value is continuous (e.g. height, time), then this can motivate treating the measurement as a quantitative dependent variable since fractional values 'make sense'. For example, even if you measure time only down to the nearest millisecond, a value of 0.5 milliseconds is possible and interpretable. In contrast, a value of 0.5 people is not.
- Are there a large number (>50) of possible values the measured variable can take? For example, a die can only take on 6 quantitative values, which is not enough.
- Are most/all of the observed values far from their bounds? Human height does not really get much smaller than about 50 cm and longer than about 220 cm, so it is technically bounded. However, in most cases, our observations are expected to not be away from these boundaries.

Logical Variables

```r
2 == 1
## [1] FALSE

"hello" == "hello"
## [1] TRUE

"hello" != "hello"
## [1] FALSE
```

```r
# are the values less than or equal to 3?
c(1,2,3,4,5,6,7,8,9,10) <= 3
##  [1]  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

```r
logical_vector = c(1,2,3,4,5,6,7,8,9,10) <= 3

as.numeric (logical_vector)
##  [1] 1 1 1 0 0 0 0 0 0 0

sum (logical_vector)
## [1] 3

sum (c(1,2,3,4,5,6,7,8,9,10) <= 3)
## [1] 3
```

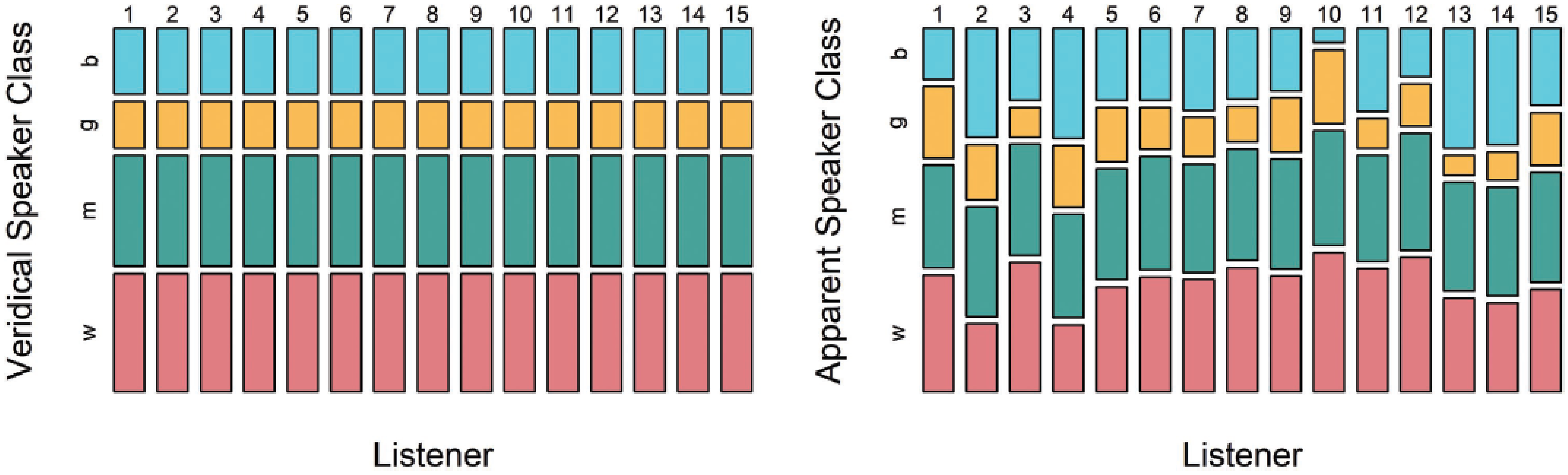
Logical Variables

```
# TRUE if f0 < 175
f0_idx = exp_data_all$f0 < 175

str (f0_idx)
##  logi [1:4170] FALSE FALSE FALSE FALSE FALSE FALSE ...

sum (f0_idx)
## [1] 1290
```

```
# get only rows where f0 < 175, i.e. where f0_idx is TRUE
low_f0 = exp_data_all[f0_idx,]

nrow(low_f0)
## [1] 1290

max(low_f0$f0)
## [1] 172
```

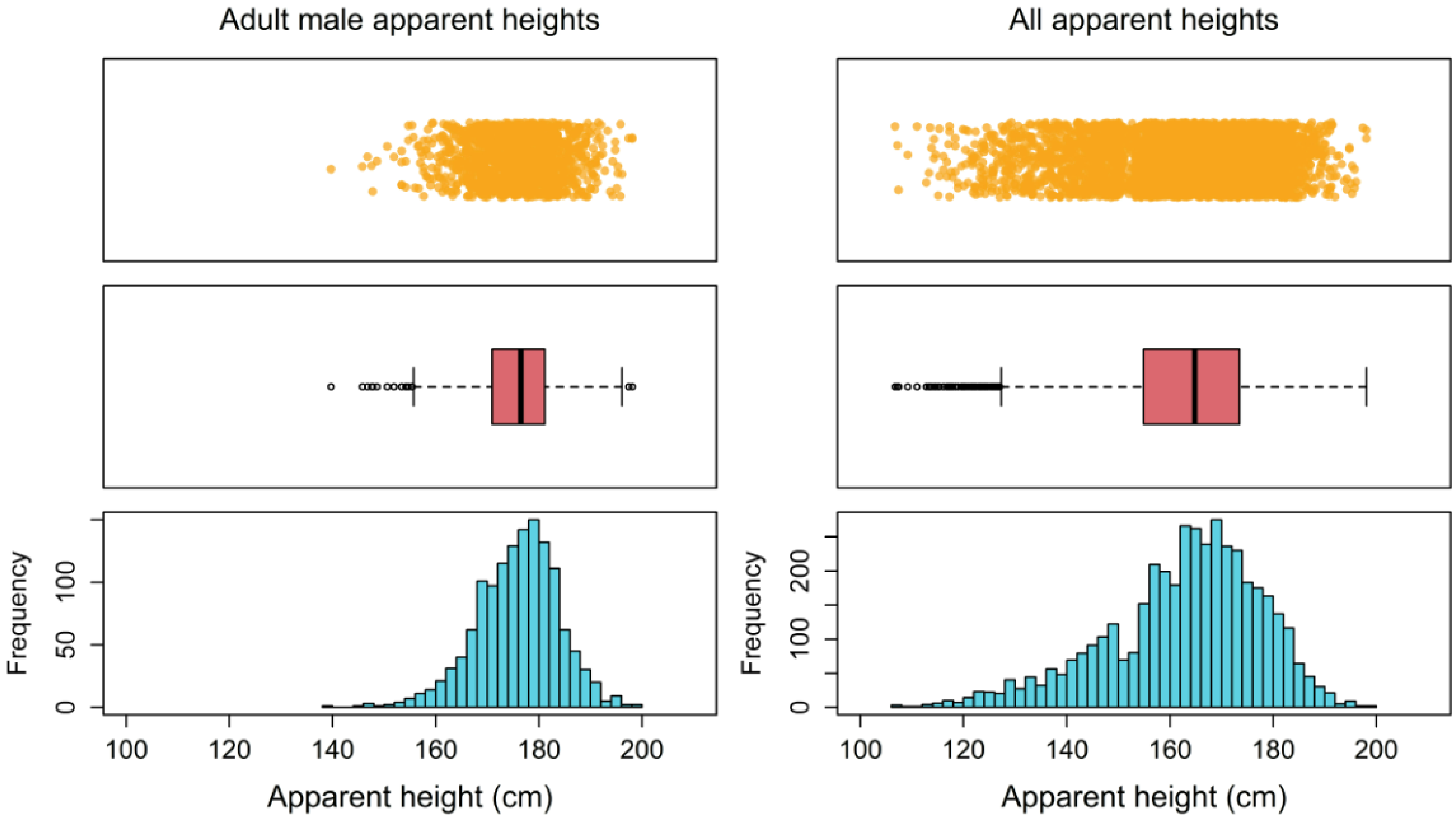
```
# get only rows where f0 >= 175, i.e. where f0_idx is FALSE
high_f0 = exp_data_all[!f0_idx,]

nrow(high_f0)
## [1] 2880

min(high_f0$f0)
## [1] 175
```

Plotting Distributions

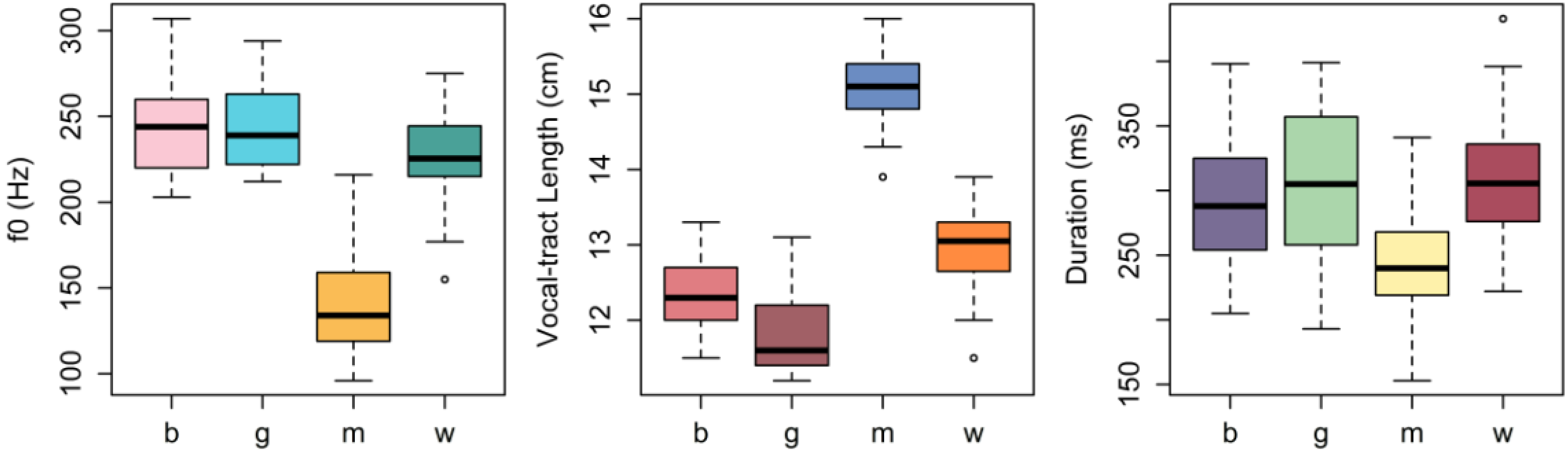



```
# table of listener and veridical speaker category
table (exp_data_all$C_v, exp_data_all$L)
##
##       1  2  3  4  5  6  7  8  9 10 11 12 13 14 15
##    b 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54
##    g 38 38 38 38 38 38 38 38 38 38 38 38 38 38 38
##    m 90 90 90 90 90 90 90 90 90 90 90 90 90 90 90
##    w 96 96 96 96 96 96 96 96 96 96 96 96 96 96 96
```

Plotting Distributions

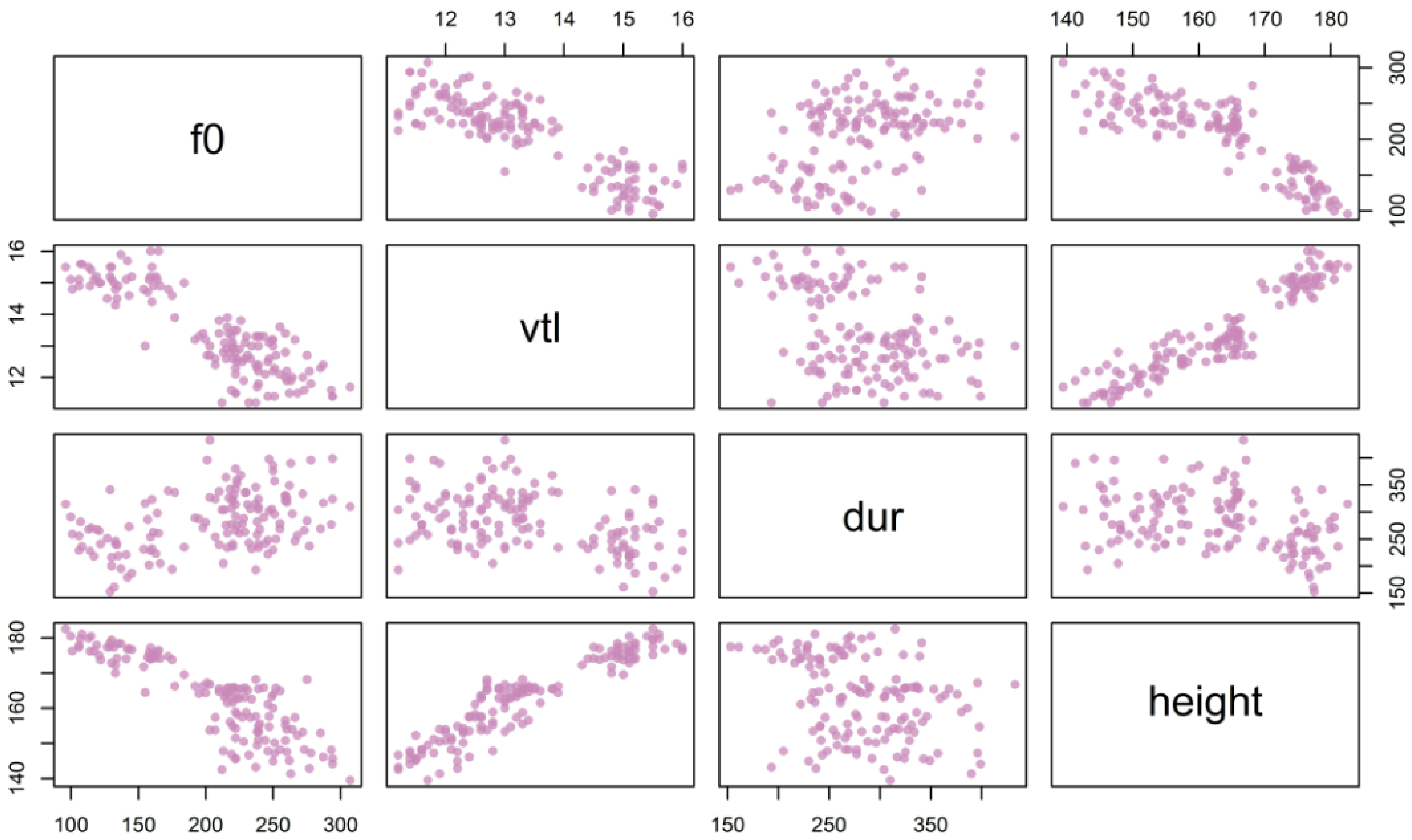


Box Plots



f0 (Hz)
100
150
200
250
300
b
g
m
w
Vocal-tract Length (cm)
12
13
14
15
16
b
g
m
w
Duration (ms)
150
250
350
b
g
m
w

Scatter Plots



Data

- The book experiment.
- Davis and Delhi temperatures.
- GDP per capita
- Male and female height

	date	meantemp	humidity	wind_spe	meanpressure
1	date	meantemp	humidity	wind_spe	meanpressure
2	1/1/2013	10	84.5	0	1015.667
3	1/2/2013	7.4	92	2.98	1017.8
4	1/3/2013	7.166667	87	4.633333	1018.667
5	1/4/2013	8.666667	71.33333	1.233333	1017.167
6	1/5/2013	6	86.83333	3.7	1016.5
7	1/6/2013	7	82.8	1.48	1018

	year	month	date	temperature
1	year	month	date	temperature
2	1950	1	1	47
3	1950	1	2	40
4	1950	1	3	30
5	1950	1	4	30
6	1950	1	5	33
7	1950	1	6	37
8	1950	1	7	37
9	1950	1	8	39
10	1950	1	9	38
11	1950	1	10	35
12	1950	1	11	41

Data

- NFL, NBA, Soccer data

A1 season_id

season_id	team_id_h	team_abb	team_nam	game_id	game_dat	matchup_	wl_home	min	fgm_home	fga_home	fg_pct_hor	fg3m_hom	fg3a_hom	fg3_pct_hc	ftm_home	fta_home	ft_pct_hon	oreb_hom	dreb_hom	reb_home	ast_home	stl_home	blk_home	tov_home	pf_home	pts_home	plus_minu	video_ava	team_id_a	team_abb	team_nam	matchup_	wl_av
21999	1.61E+09	VAN	Vancouve	29900013	########	VAN vs. PC	L	240	30	75	0.4	2	7	0.286	24	34	0.706	15	26	41	19	4	3	17	23	86	-20	0	1.61E+09	POR	Portland T	POR @ VA	W
21999	1.61E+09	CHH	Charlotte I	29900005	########	CHH vs. OI	W	240	34	82	0.415	10	22	0.455	22	37	0.595	14	33	47	25	13	5	21	25	100	14	0	1.61E+09	ORL	Orlando M	ORL @ CHI	L
21999	1.61E+09	NJN	New Jerse	29900002	########	NJN vs. IN	L	240	36	81	0.444	5	12	0.417	35	46	0.761	15	30	45	11	11	4	18	39	112	-7	0	1.61E+09	IND	Indiana P	IND @ NJN	W
21999	1.61E+09	LAC	Los Angele	29900012	########	LAC vs. SE	L	240	35	78	0.449	6	16	0.375	16	27	0.593	12	33	45	14	4	9	18	24	92	-12	0	1.61E+09	SEA	Seattle Su	SEA @ LAC	W
21999	1.61E+09	WAS	Washingt	29900003	########	WAS vs. AT	W	240	39	88	0.443	3	10	0.3	13	16	0.813	12	30	42	23	5	5	15	30	94	7	0	1.61E+09	ATL	Atlanta Ha	ATL @ WA:	L
21999	1.61E+09	TOR	Toronto R	29900006	########	TOR vs. BC	L	240	33	75	0.44	3	13	0.231	21	30	0.7	10	26	36	22	9	7	17	20	90	-13	0	1.61E+09	BOS	Boston Ce	BOS @ TO	W
21999	1.61E+09	SAS	San Anton	29900009	########	SAS vs. PH	W	240	31	72	0.431	6	10	0.6	21	29	0.724	15	35	50	22	4	12	22	21	89	13	0	1.61E+09	PHI	Philadelpl	PHI @ SAS	L
21999	1.61E+09	UTA	Utah Jazz	29900011	########	UTA vs. LA	L	240	29	67	0.433	3	10	0.3	23	28	0.821	8	30	38	20	7	6	18	26	84	-7	0	1.61E+09	LAL	Los Angele	LAL @ UTA	W
21999	1.61E+09	NYK	New York	29900001	########	NYK vs. CL	W	240	32	74	0.432	6	15	0.4	22	27	0.815	9	35	44	17	10	10	18	23	92	8	0	1.61E+09	CLE	Cleveland	CLE @ NYK	L
21999	1.61E+09	MIA	Miami Hea	29900004	########	MIA vs. DE	W	290	41	106	0.387	10	29	0.345	36	48	0.75	22	44	66	24	7	5	15	39	128	6	0	1.61E+09	DET	Detroit Pis	DET @ MIA	L
21999	1.61E+09	DEN	Denver Nu	29900010	########	DEN vs. PH	W	265	40	89	0.449	8	18	0.444	19	26	0.731	13	42	55	25	6	16	19	17	107	5	0	1.61E+09	PHX	Phoenix S	PHX @ DEI	L

Rk	Player	Tm	Age	Pos	G	GS	Att	Yds	TD	X1D	Succ.	Lng	Y.A	Y.G	Fmb	code
1	DeMarco I	DAL	26	RB	16	16	392	1845	13	89	53.1	51	4.7	115.3	5	MurrDe00
2	LeSean Mc	PHI	26	RB	16	16	312	1319	5	68	44.6	53	4.2	82.4	4	McCoLe01
3	Le'Veon Be	PIT	22	RB	16	16	290	1361	8	75	49.3	81	4.7	85.1	0	BellLe00
4	Marshawr	SEA	28	RB	16	14	280	1306	13	67	52.1	79	4.7	81.6	3	LyncMa00
5	Matt Forte	CHI	29	RB	16	16	266	1038	6	65	48.9	32	3.9	64.9	2	FortMa00
6	Alfred Mor	WAS	26	RB	16	16	265	1074	8	53	46	30	4.1	67.1	2	MorrAl00
7	Arian Fost	HOU	28	RB	13	13	260	1246	8	55	45	51	4.8	95.8	2	FostAr00
8	Frank Gor	SFO	31	RB	16	16	255	1106	4	55	48.2	52	4.3	69.1	2	GoreFr00
9	Eddie Lacy	GNB	24	RB	16	16	246	1139	9	60	45.9	44	4.6	71.2	3	LacyEd00
10	Justin For:	BAL	29	RB	16	14	235	1266	8	57	42.6	52	5.4	79.1	1	ForsJu00
11	Mark Ingra	NOR	25	RB	13	9	226	964	9	52	49.1	31	4.3	74.2	3	IngrMa01

Assignment 1

- Write a report using a qmd file that investigates:

1. Apparent height, f0, or VTL (vocal-tract length) in the book data.

2. Or investigate any data you want that is ‘appropriate’ for the next few classes (see next slide).

- Submit the report and the qmd file.

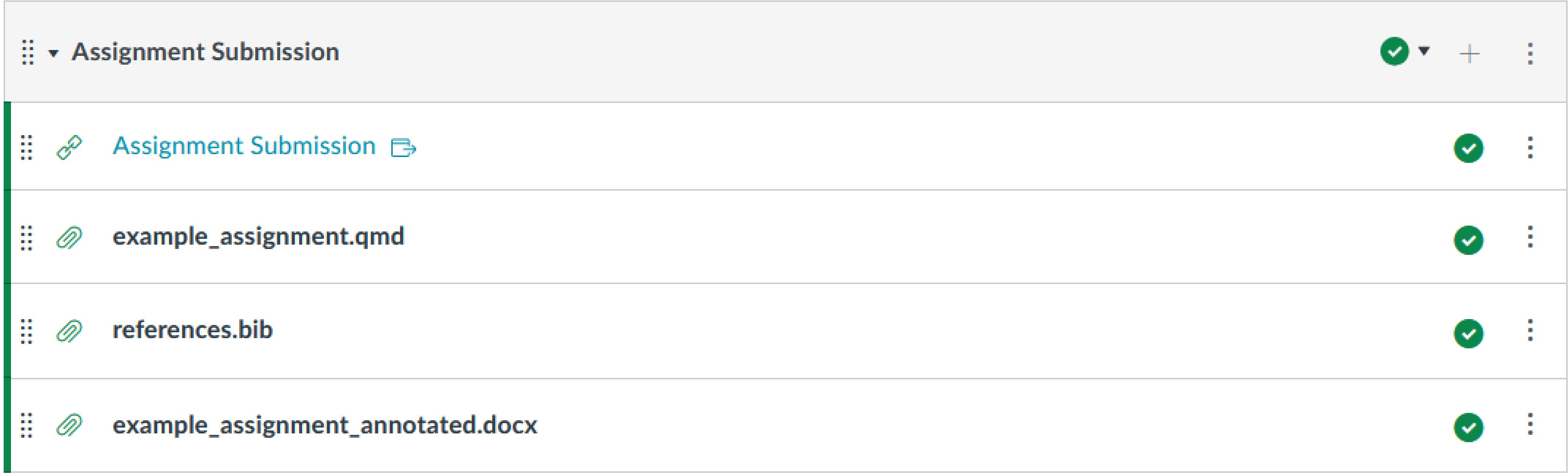
Assignment 1

- Use the paper to establish that the variable you are investigating is:

  1. Quantitative
  2. Is not highly skewed
  3. Values are not close to its upper or lower bounds

- Include at least two plots that corroborate this and describe the information they represent in the figure legend.

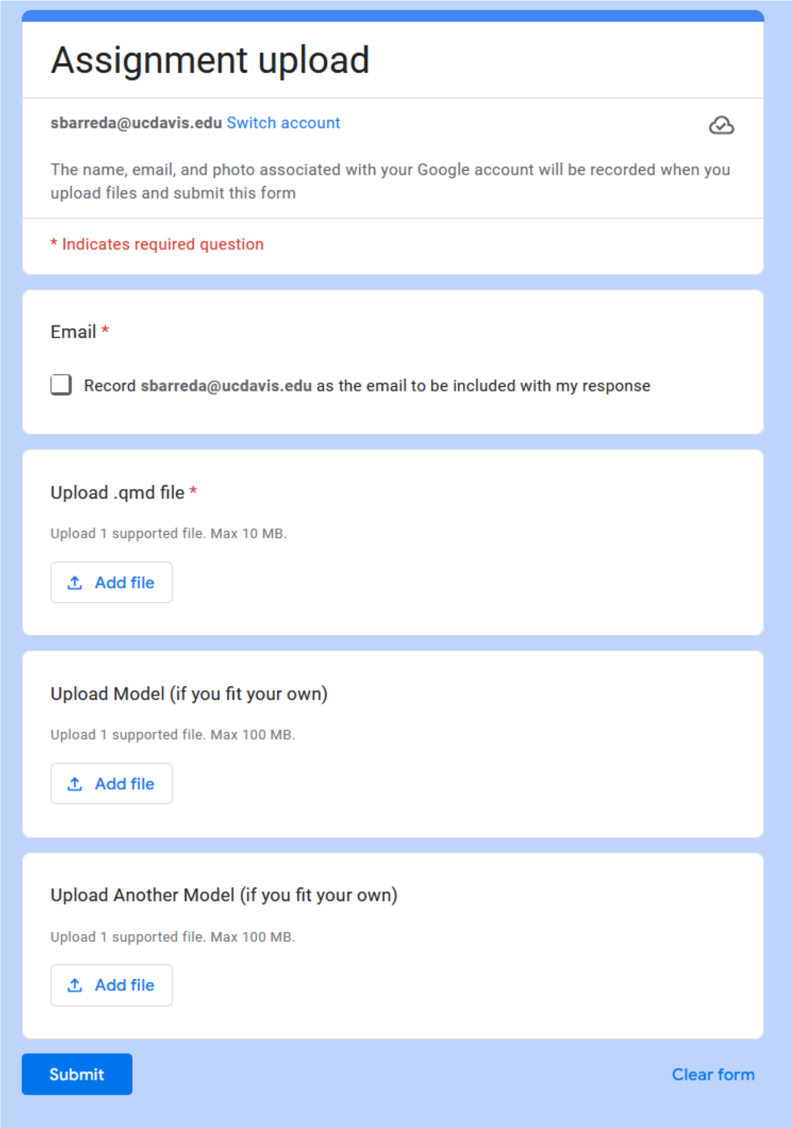
Assignment Submission



.Rmd Files

- Two options:
  - Use the ‘minimal_bookdown_example.zip’ example.
  - Use the default new Rmd file and ‘knit’ it.

Assignment Submission


Assignment upload

sbarreda@ucdavis.edu Switch account

The name, email, and photo associated with your Google account will be recorded when you upload files and submit this form

* Indicates required question

Email *

Record sbarreda@ucdavis.edu as the email to be included with my response

Upload .qmd file *

Upload 1 supported file. Max 10 MB.

Add file

Upload Model (if you fit your own)

Upload 1 supported file. Max 100 MB.

Add file

Upload Another Model (if you fit your own)

Upload 1 supported file. Max 100 MB.

Add file

Submit

Clear form

- Please submit the qmd files and model (if you fit one) using this form.
- Submit the word or pdf document version on Canvas.