ADDITIONAL EXERCISES BY CHAPTER

IMPORTANT: Feel free to draw on these materials when using our book as the main textbook in your course, but please do not share the additional real-world datasets provided or the solutions to these exercises with anyone outside your class. For example, please do not post them on a non-gated website or in a public GitHub repository.

The contents of this folder are exclusively for instructors using Llaudet and Imai, *Data Analysis for Social Science: A Friendly and Practical Introduction* (Princeton University Press, 2022) as the main textbook in their course. It contains additional exercises organized by chapter, which instructors can use as in-class exercises or as take-home problem sets. The additional exercises provided ask similar questions as the exercises in the book, but they use different real-world datasets so that students get to see the same material in different contexts. Below is a quick overview of the exercises included:

Chapter 1: Introduction

- Estimating the Bias in Self-Reported Turnout - Part I: Loading and Making Sense of Data
  - Research Question: Is There Bias in Self-Reported Turnout?
  - Skills Practiced: Load and make sense of data.
  - Code Used: read.csv(), <-, head(), dim()
  - Difficulty: Easy

- Estimating the Bias in Self-Reported Turnout - Part II: Computing and Interpreting Means
  - Research Question: Is There Bias in Self-Reported Turnout?
  - Skills Practiced: Use arithmetic operators to create new variables from existing ones, plus compute and interpret means.
  - Code Used: read.csv(), <-, head(), /, *, $, mean()
  - Difficulty: Intermediate (if assigned in Chapter 1), Easy (if assigned in Chapter 3)

- Effects of A Criminal Record in Labor Market - Part I: Loading and Making Sense of Data
  - Research Question: Does Having a Criminal Record Affect the Chances of Receiving A Call Back for a Job Interview?
  - Skills Practiced: Load and make sense of data.
  - Code Used: read.csv(), <-, head(), dim()
  - Based on: Devah Pager. 2003. "The Mark of a Criminal Record." *American Journal of Sociology*, 108 (5): 937-75.
  - Difficulty: Easy

- Effects of A Criminal Record in Labor Market - Part II: Computing and Interpreting Means
  - Research Question: Does Having a Criminal Record Affect the Chances of Receiving A Call Back for a Job Interview? (Focus on White Applicants)
  - Skills Practiced: Compute and interpret means, among other things. (The last few questions refer to Chapter 2 of DSS)

  - Code Used: read.csv(), <-, head(), mean(), $
  - Based on: Devah Pager. 2003. "The Mark of a Criminal Record." *American Journal of Sociology*, 108 (5): 937-75.
  - Difficulty: Easy

- Effects of Having a Female Politician in Rural India - Part I: Loading and Making Sense of Data
  - Research Question: Do Women Promote Different Policies than Men?
  - Skills Practiced: Load and make sense of data.
  - Code Used: read.csv(), <-, head(), dim()
  - Based on: Raghabendra Chattopadhyay and Esther Duflo. 2004. "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India." *Econometrica*, 72 (5): 1409-43.
  - Difficulty: Easy

- Effects of Having a Female Politician in Rural India - Part II: Computing and Interpreting Means
  - Research Question: Do Women Promote Different Policies than Men?
  - Skills Practiced: Compute and interpret means, among other things. (The last few questions refer to Chapter 2 of DSS)
  - Code Used: read.csv(), <-, head(), mean(), $
  - Based on: Raghabendra Chattopadhyay and Esther Duflo. 2004. "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India." *Econometrica*, 72 (5): 1409-43.
  - Difficulty: Easy

- Effect of Social Pressure Message on Probability of Voting - Part I: Loading and Making Sense of Data, and Computing and Interpreting Means
  - Research Question: Does Social Pressure Affect Turnout?
  - Skills Practiced: Load and make sense of data, and compute and interpret means.
  - Code Used: read.csv(), <-, head(), dim(), mean(), $
  - Based on: Alan Gerber, Donald Green, and Christopher Larimer. 2008. "Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment." *American Political Science Review*, 102(1): 33-48.
  - Difficulty: Easy-Intermediate

Chapter 2: Estimating Causal Effects with Randomized Experiments

- Effects of A Criminal Record in Labor Market - Part II: Computing and Interpreting Means
  - Research Question: Does Having a Criminal Record Affect the Chances of Receiving A Call Back for a Job Interview? (Focus on White Applicants)
  - Skills Practiced: Identify treatment and outcome variables and treatment and control groups, among other things.
  - Code Used: read.csv(), <-, head(), mean(), $
  - Based on: Devah Pager. 2003. "The Mark of a Criminal Record." *American Journal of Sociology*, 108 (5): 937-75. Difficulty: Easy

- Effects of A Criminal Record in Labor Market - Part III: Estimating an Average Causal Effect
  - Research Question: Does Having a Criminal Record Affect the Chances of Receiving A Call Back for a Job Interview? (Focus on White Applicants)

  - Skills Practiced: Estimate an average treatment effect using data from a randomized experiment and the difference-in-means estimator, and write a conclusion statement.
  - Code Used: read.csv(), <-, head(), mean(), $, [], ==
  - Based on: Devah Pager. 2003. "The Mark of a Criminal Record." *American Journal of Sociology*, 108 (5): 937-75.
  - Difficulty: Easy

- Effects of Having a Female Politician in Rural India - Part II: Computing and Interpreting Means
  - Research Question: Do Women Promote Different Policies than Men?
  - Skills Practiced: Identify treatment and outcome variables and treatment and control groups, among other things.
  - Code Used: read.csv(), <-, head(), mean(), $
  - Based on: Raghabendra Chattopadhyay and Esther Duflo. 2004. "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India." *Econometrica*, 72 (5): 1409-43.
  - Difficulty: Easy

- Effects of Having a Female Politician in Rural India - Part III: Estimating an Average Causal Effect
  - Research Question: Do Women Promote Different Policies than Men?
  - Skills Practiced: Estimate an average treatment effect using data from a randomized experiment and the difference-in-means estimator, and write a conclusion statement.
  - Code Used: read.csv(), <-, head(), mean(), $, [], ==
  - Based on: Raghabendra Chattopadhyay and Esther Duflo. 2004. "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India." *Econometrica*, 72 (5): 1409-43.
  - Difficulty: Easy

- Effect of Social Pressure Message on Probability of Voting - Part II: Estimating an Average Causal Effect
  - Research Question: Does Social Pressure Affect Turnout?
  - Skills Practiced: Create new variables, estimate average treatment effects using data from a randomized experiment and the difference-in-means estimator, and write conclusion statements.
  - Code Used: read.csv(), <-, head(), ifelse(), $, ==, mean(), []
  - Based on: Alan Gerber, Donald Green, and Christopher Larimer. 2008. "Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment." *American Political Science Review*, 102(1): 33-48.
  - Difficulty: Easy-Intermediate

Chapter 3: Inferring Population Characteristics via Survey Research

- Estimating the Bias in Self-Reported Turnout - Part II: Computing and Interpreting Means
  - Research Question: Is There Bias in Self-Reported Turnout?
  - Skills Practiced: Use arithmetic operators to create new variables from existing ones, plus compute and interpret means.
  - Code Used: read.csv(), <-, head(), /, *, $, mean()
  - Difficulty: Intermediate (if assigned in Chapter 1), Easy (if assigned in Chapter 3)

- Estimating the Bias in Self-Reported Turnout - Part III: Subsetting Variables and Creating Histograms
  - Research Question: Is There Bias in Self-Reported Turnout?
  - Skills Practiced: Create new variables, visualize the distribution of a variable, subset variables, and compute and interpret means.
  - Code Used: read.csv(), <-, head(), /, *, $, hist(), [], mean()
  - Difficulty: Intermediate

- Evidence of Data Fabrication
  - Research Question: Do We Find Evidence of Data Fabrication?
  - Skills Practiced: Create histograms, compute descriptive statistics, create scatter plots, compute correlation coefficients, and use optional arguments in R functions.
  - Code Used: read.csv(), <-, head(), dim(), $, hist(), mean(), median(), sd(), plot(), cor()
  - Based on: David Broockman, Joshua Kalla, and Peter Aronow. "Irregularities in LaCour (2014)"
  - Difficulty: Intermediate-Advanced

- Effects of Having a Female Politician in Rural India - Part IV: Visualizations and Correlations
  - Research Question: Do Women Promote Different Policies than Men?
  - Skills Practiced: Create and make sense of visualizations and compute and interpret the correlation between two variables, among other things.
  - Code Used: read.csv(), <-, head(), hist(), $, plot(), cor()
  - Based on: Raghabendra Chattopadhyay and Esther Duflo. 2004. "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India." *Econometrica*, 72 (5): 1409-43.
  - Difficulty: Easy

- Effect of Assassination of Leaders on Level of Democracy - Part I: Visualizations and Correlations
  - Research Question: What is the Effect of the Death of the Leader on the Level of Democracy?
  - Skills Practiced: Computing the number of observations in a dataset, create table of frequencies, compute and interpret means, create density histograms, subset variables, create scatter plots, and compute correlation coefficients.
  - Code Used: read.csv(), <-, head(), dim(), table(), $, mean(), hist(), [], ==, plot(), cor()
  - Based on: Benjamin Jones and Benjamin Olken. 2009. "Hit or Miss? The Effect of Assassinations on Institutions and War." *American Economic Journal: Macroeconomics*, 1 (2): 55-87.
  - Difficulty: Easy

Chapter 4: Predicting Outcome Using Linear Regression

- Predicting Course Grades - Part I: Predicting Final Exam Scores
  - Goal: Predict Final Exam Scores Using Midterm Scores (Note: The outcome is non-binary)
  - Skills Practiced: Fit a line to make predictions when Y is non-binary, including computing correlations, creating scatter plots, adding the fitted line to the plot, and computing R-squared.

  - Code Used: read.csv(), <-, head(), cor(), $, lm(), plot(), abline()
  - Difficulty: Easy
  - Note: We recommend using real, historical performance data from your own class.

- Predicting Course Grades - Part II: Predicting Overall Scores
  - Goal: Predict Overall Scores in the Course Using Midterm Scores (Note: The outcome is non-binary)
  - Skills Practiced: Fit a line to make predictions when Y is non-binary, including computing correlations, creating scatter plots, adding the fitted line to the plot, and computing R-squared.
  - Code Used: read.csv(), <-, head(), cor(), $, lm(), plot(), abline()
  - Difficulty: Easy, especially if assigned after Part I
  - Note: We recommend using real, historical performance data from your own class.

- Predicting Course Grades - Part III: Predicting Probability of Earning an A or A-
  - Goal: Predict Probability of Earning an A or A- in the Course Using Midterm Scores (Note: The outcome is binary)
  - Skills Practiced: Fit a line to make predictions when Y is binary, including computing correlations, creating scatter plots, adding the fitted line to the plot, and computing R-squared.
  - Code Used: read.csv(), <-, head(), cor(), $, lm(), plot(), abline()
  - Difficulty: Easy, especially if assigned after Parts I and II
  - Note: We recommend using real, historical performance data from your own class.

- Predicting Elections Using Betting Markets
  - Goal: Predict State-Level Electoral Outcomes Based on Closing Price of Contracts in Betting Markets (Note: The outcome is non-binary)
  - Skills Practiced: Create new variables, create histograms, compute correlations, create scatter plots, fit a linear model for predictive purposes, make predictions using the fitted line, compute and interpret R-squared.
  - Code Used: read.csv(), <-, head(), dim(), $, hist(), cor(), lm(), plot(), abline()
  - Difficulty: Intermediate

Chapter 5: Estimating Causal Effects with Observational Data

- Effect of Black Candidates on Black Turnout
  - Research Question: Does Having a Black Candidate Running Increase Black Turnout?
  - Skills Practiced: Estimating causal effects with observational data by fitting a multiple linear regression model where Y is the outcome variable of interest, X1 is the treatment variable, and X2 is the only confounding variable.
  - Code Used: read.csv(), <-, head(), mean(), $, [], ==, lm(), cor()
  - Based on: Bernard Fraga. 2016. "Candidates or Districts? Reevaluating the Role of Race in Voter Turnout." *American Journal of Political Science*, 60: 97-122.
  - Difficulty: Easy-Intermediate

- Effect of Assassination of Leaders on Level of Democracy - Part II: Fitting a Line to Compute the Difference-in-Means Estimator
  - Research Question: What is the Effect of the Death of the Leader on the Level of Democracy?
  - Skills Practiced: Fit a linear model to compute the difference-in-means estimator.
  - Code Used: read.csv(), <-, head(), mean(), $, [], ==, lm()
  - Based on: Benjamin Jones and Benjamin Olken. 2009. "Hit or Miss? The Effect of Assassinations on Institutions and War." *American Economic Journal: Macroeconomics*, 1 (2): 55-87.
  - Difficulty: Easy

- Effect of Assassination of Leaders on Level of Democracy - Part III: Controlling for Confounders
  - Research Question: What is the Effect of the Death of the Leader on the Level of Democracy?
  - Skills Practiced: Estimating causal effects with observational data by fitting a multiple linear regression model where Y is the outcome variable of interest, X1 is the treatment variable, and X2 is the only confounding variable.
  - Code Used: read.csv(), <-, head(), $, [], ==, lm()
  - Based on: Benjamin Jones and Benjamin Olken. 2009. "Hit or Miss? The Effect of Assassinations on Institutions and War." *American Economic Journal: Macroeconomics*, 1 (2): 55-87.
  - Difficulty: Easy-Intermediate

- Effect of Political TV Ads on Turnout
  - Research Question: Do Negative Political TV Ads Decrease Voter Turnout?
  - Skills Practiced: Summarize and evaluate a published social scientific study, including discuss internal and external validity.
  - Code Used: None
  - Based on: Ansolabehere, Stephen, Shanto Iyengar, Adam Simon and Nicholas Valentino. 1994. "Does Attack Advertising Demobilize the Electorate?" *The American Political Science Review*, Vol. 88, No. 4, pp. 829-838.
  - Difficulty: Easy

Chapter 6: Probability

- Multiple Choice Questions on Probability
  - Goal: Review concepts discussed in Chapter 6
  - Code Used: None
  - Difficulty: Easy

Chapter 7: Quantifying Uncertainty

- Effects of A Criminal Record in Labor Market - Part IV: Focus on White Applicants
  - Research Question: Does Having a Criminal Record Affect the Chances of Receiving A Call Back for a Job Interview? (Focus on White Applicants, binary outcome)

  - Skills Practiced: Estimate an average treatment effect using data from a randomized experiment and a linear model, and determine whether the estimated average treatment effect is statistically significant at the 5% level.
  - Code Used: read.csv(), <-, head(), lm(), $, summary()$coeff
  - Based on: Devah Pager. 2003. "The Mark of a Criminal Record." *American Journal of Sociology*, 108 (5): 937-75.
  - Difficulty: Easy

- Effects of A Criminal Record in Labor Market - Part V: Focus on Black Applicants
  - Research Question: Does Having a Criminal Record Affect the Chances of Receiving A Call Back for a Job Interview? (Focus on Black Applicants, binary outcome)
  - Skills Practiced: Answer (1) What is the estimated average treatment effect? (2) Is the effect statistically significant at the 5% level? (3) Can we interpret the effect as causal? (4) Can we generalize the results?
  - Code Used: read.csv(), <-, head(), lm(), $, summary()$coeff
  - Based on: Devah Pager. 2003. "The Mark of a Criminal Record." *American Journal of Sociology*, 108 (5): 937-75.
  - Difficulty: Easy

- Effects of Having a Female Politician in Rural India - Part V: Effect on Drinking Water Facilities
  - Research Question: Do Women Promote Different Policies than Men? Effect of Having a Female Politician on Drinking Water Facilities in Rural India (Note: The outcome is non-binary)
  - Skills Practiced: Estimate average treatment effects using data from a randomized experiment and a linear model, determine whether the estimated average treatment effect is statistically significant at the 5% level, and discuss the internal and external validity of the study. (Note: The effect here is statistically significant.)
  - Code Used: read.csv(), <-, head(), lm(), $, summary()$coeff
  - Based on: Raghabendra Chattopadhyay and Esther Duflo. 2004. "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India." *Econometrica*, 72 (5): 1409-43.
  - Difficulty: Easy

- Effects of Having a Female Politician in Rural India - Part VI: Effect on Irrigation Facilities
  - Research Question: Do Women Promote Different Policies than Men? Effect of Having a Female Politician on Irrigation Facilities in Rural India (Note: The outcome is non-binary)
  - Skills Practiced: Estimate average treatment effects using data from a randomized experiment and a linear model, determine whether the estimated average treatment effect is statistically significant at the 5% level, and discuss the internal and external validity of the study. (Note: The effect here is NOT statistically significant.)
  - Code Used: read.csv(), <-, head(), lm(), $, summary()$coeff
  - Based on: Raghabendra Chattopadhyay and Esther Duflo. 2004. "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India." *Econometrica*, 72 (5): 1409-43.
  - Difficulty: Easy

- Predicting Course Grades - Part IV: Quantifying Uncertainty
  - Goal: Using Midterm Scores, Predict Final Exam Scores, Overall Course Scores, and Probability of Earning an A or A- in the Course, and Construct 95% Confidence Intervals for the Predictions. (Note: Two of the outcomes are non-binary, the other is binary.)
  - Skills Practiced: Fit a line to make predictions, create scatter plots, add the fitted line to the scatter plot, and construct 95% confidence intervals for our predictions
  - Code Used: read.csv(), <-, head(), $, lm(), plot(), abline(), predict()
  - Difficulty: Easy
  - Note: We recommend using real, historical performance data from your own class.

- Effects of Black-Sounding Names on Call Backs for Job Interviews
  - Research Question: Is There Racial Discrimination in the Labor Market? (Note: The outcome is binary)
  - Skills Practiced: Answer (1) What is the estimated average treatment effect? (2) Is the effect statistically significant at the 5% level? (3) Can we interpret the effect as causal? (4) Can we generalize the results?
  - Code Used: read.csv(), <-, head(), hist(), $, ifelse(), ==, lm(), plot(), abline()
  - Based on: Marianne Bertrand and Sendhil Mullainathan. 2004. "Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination." *American Economic Review*, 94 (4): 991-1013.
  - Difficulty: Easy

- Effect of Small Classes on Student Outcomes
  - Research Question: Do Small Classes in Elementary School Improve Student Outcomes? (Note: Two of the outcomes are non-binary, the other is binary.)
  - Skills Practiced: Estimate average treatment effects using data from a randomized experiment and a linear model, and determine whether the estimated average treatment effects are statistically significant at the 5% level.
  - Code Used: read.csv(), <-, head(), ifelse(), $, ==, lm(), plot(), abline(), summary()$coeff
  - Based on: Frederick Mosteller, "The Tennessee Study of Class Size in the Early School Grades," *Future of Children* 5, no.2 (1995): 113-27.
  - Difficulty: Easy

- Effect of Social Pressure Message on Probability of Voting - Part III: Estimate an Average Causal Effect, Determine Statistical Significance, and Discuss Internal and External Validity
  - Research Question: Does Social Pressure Affect Turnout? (Note: The outcome is binary)
  - Skills Practiced: Estimate average treatment effects using data from a randomized experiment and a linear model, determine whether the estimated average treatment effect is statistically significant at the 5% level, and discuss the internal and external validity of the study.
  - Code Used: read.csv(), <-, head(), ifelse(), $, ==, lm(), summary()$coeff
  - Based on: Alan Gerber, Donald Green, and Christopher Larimer. 2008. "Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment." *American Political Science Review*, 102(1): 33-48.
  - Difficulty: Easy-Intermediate

NOTE #1: The PDFs enclosed include the exercises' solutions. The PDFs with the exercises without the solutions (which you may want to share with your students), are inside the folder with the name of the exercise + files.

NOTE #2: If you would like to make any changes to the exercises, you will need to modify the corresponding Rmarkdown files. To do so:

- open the corresponding .Rmd file in RStudio (the .Rmd file is inside the folder with the name of the exercise + files)
- make the changes
- knit the file (by clicking on the "Knit" icon or by using the shortcut shift+command+k in Mac or shift+ctrl+k in Windows)

Inside the folders that contain the Rmarkdown files, there are other files that are necessary for the .Rmd file to knit (e.g., datasets and preambles). This means that you should not move the .Rmd files to a different location. In other words, you should knit them inside the folders provided. After knitting an .Rmd file, a new PDF will appear in the folder. By default, the PDF will contain the solutions to the questions. To produce the PDF without the solutions, delete the 1 in line 10 of the corresponding .Rmd file. In other words, delete the 1 in this piece of code:

```
\newcommand{\solutions}{1}
```

If you have any feedback and/or questions, please do not hesitate to contact me at ellaudet@gmail.com.