Table of Contents

Content			Page
Overview and Philosophy			9
Scope and Sequence			15
UNIT 1	**Campaign**	**Topics**	
Daily Overview			21
Essential Concepts			22
Section 1: Data are all Around			24
Lesson 1: Data Trails		Defining data, consumer privacy	26
Lesson 2: Stick Figures		Organizing & collecting data	28
Lesson 3: Data Structures		Organizing data, rows & columns, variables	31
Lesson 4: The Data Cycle		Data cycle, statistical questions	34
Lesson 5: So Many Questions		Statistical questions, variability	38
Lesson 6: What Do I Eat?	Food Habits	Collecting data, statistical questions	40
Lesson 7: Setting the Stage	Food Habits – data	Participatory sensing	43
Section 2: Visualizing Data			47
Lesson 8: Tangible Plots	Food Habits – data	Dotplots, minimum/maximum, frequency	49
Lesson 9: What is Typical?	Food Habits – data	Typical value, center	52
Lesson 10: Making Histograms	Food Habits – data	Histograms, bin widths	55
Lesson 11: What Shape Are You In?	Food Habits – data	Shape, center, spread	58
Lesson 12: Exploring Food Habits	Food Habits – data	Single & multi-variable plots,	60
Lesson 13: Rstudio Basics	Food Habits – data	Intro to RStudio	62
Lab 1A: Data, Code & Rstudio	Food Habits – data	RStudio basics	65
Lab 1B: Get the Picture?	Food Habits – data	Variable types, bar graphs, histograms	68
Lab 1C: Export, Upload, Import	Food Habits – data	Importing data	71
Lesson 14: Variables, Variables, Variables		Multi-variable plots	75
Lab 1D: Zooming Through Data		Subsetting	79
Lab 1E: What's the Relationship?		Multi-variable plots	82
Practicum: The Data Cycle & My Food Habits	Food Habits	Data cycle, variability	85
Section 3: Would You Look at the Time			87
Lesson 15: Americans' Time on Task	Time Use – data	Evaluating claims	89
Lab 1F: A Diamond in the Rough	Time Use - data	Cleaning names, categories, and strings	94
Lesson 16: Categorical Associations	Time Use - data	Joint relative frequencies in 2-way tables	98
Lesson 17: Interpreting 2-Way Tables	Time Use - data	Marginal & conditional relative frequencies	100
Lab 1G: What's the FREQ?	Time Use – data	2-way tables, tally	103
Lab 1H: Our Time	Time Use	Data cycle, synthesis	105
Practicum: Teen Depression		Statistical questions, interpreting plots	106
End of Unit Project and Oral Presentation: Analyzing Data to Evaluate Claims		Data cycle	108

Table of Contents (continued)

UNIT 2	Campaign	Topics	Page
Daily Overview			110
Essential Concepts			111
Section 1: What is Your True Color?			113
Lesson 1: What Is Your True Color?	Personality Color - data	Subsets, relative frequency	115
Lesson 2: What Does *Mean* Mean?	Personality Color	Measures of center – mean	118
Lesson 3: Median in the Middle	Personality Color	Measures of center – median	122
Lesson 4: How Far is it From Typical?	Personality Color	Measures of spread – MAD	126
Lab 2A: All About Distributions	Personality Color	Measures of center & spread – mean, median, MAD	130
Lesson 5: Human Boxplots		Boxplots, IQR	132
Lesson 6: Face Off		Comparing distributions	136
Lesson 7: Plot Match		Comparing distributions	139
Lab 2B: Oh the Summaries …	Personality Color	Boxplots, IQR, numerical summaries, custom functions	142
Practicum: The Summaries	Food Habits or Time Use	Statistical questions, comparing distributions	145
Section 2: How Likely is it?			147
Lesson 8: How Likely is it?		Probability, simulations	149
Lesson 9: Bias Detective		Simulations to detect bias	152
Lesson 10: Marbles, Marbles		Probability, with replacement	155
Lab 2C: Which Song Plays Next?		Probability of simple events, do loops, set.seed()	157
Lesson 11: This AND/OR That		Compound probabilities	160
Lab 2D: Queue it Up!		Probability with & without replacement, sample()	163
Practicum: Win, Win, Win		Probability estimation through repeated simulations	166
Section 3: Are You Stressing or Chilling?			167
Lesson 12: Don't Take My Stress Away	Stress/Chill – data	Introduction to campaign	169
Lesson 13: The Horror Movie Shuffle	Stress/Chill – data	Chance differences – cat var	173
Lab 2E: The Horror Movie Shuffle	Stress/Chill – data	Inference for categorical variable, do loops, shuffle()	177
Lesson 14: The Titanic Shuffle	Stress/Chill – data	Chance differences – num var	180
Lab 2F: The Titanic Shuffle	Stress/Chill – data	Inference for numerical variable, do loops, shuffle()	184
Lesson 15: Tangible Data Merging	Stress/Chill – data	Merging datasets	186
Lab 2G: Getting it Together	Stress/Chill & Personality Color	Merging datasets, stacking vs. joining	189
Practicum: What Stresses Us?	Stress/Chill & Personality Color	Answering statistical questions of merged data	191
Section 4: What's Normal?			192
Lesson 16: What is Normal?		Introduction to normal curve	194
Lesson 17: Normal Measure of Spread		Measures of spread - SD	198
Lesson 18: Shuffling with Normal		z-scores, shuffling	201
Lab 2H: Eyeballing Normal		Normal curves overlaid on distributions & simulated data	204
Lab 2I: R's Normal Distribution Alphabet		Normal probability, rnorm(), pnorm(), quantiles, qnorm()	206
End of Unit Project: Asking and Answering Statistical Questions of Our Own Data	Stress/Chill, Persnality Color, Habits, or Time Use	Synthesis of above	208

Table of Contents (continued)

Unit 3	Campaign	Topics	Page
Daily Overview			210
Essential Concepts			211
Section 1: Testing, Testing…1, 2, 3…			213
Lesson 1: Anecdotes vs. Data		Reading articles critically, data	215
Lesson 2: What is an Experiment?		Experiments, causation	218
Lesson 3:Let's Try an Experiment!		Random assignments, confounding factors	221
Lesson 4: Predictions, Predictions		Visualizations, predictions	223
Lesson 5: Time Perception Experiment		Elements of an experiment	225
Lab 3A: The results are in!			227
Practicum: Music to my Ears		Design an experiment	228
Section 2: Would You Look at That?			229
Lesson 6: Observational Studies		Observational study	231
Lesson 7: Observational Studies vs. Experiments		Observational study, experiment	234
Lesson 8: Monsters that Hide in Observational Studies		Observational study, confounding factors	236
Lab 3B: Confound it all!		Confounding factors	240
Section 3: Are You Asking Me?			242
Lesson 9: Survey Says…		Survey	244
Lesson 10: We're So Random		Data collection, Random samples	248
Lesson 11: The Gettysburg Address		Sampling Bias	252
Lab 3C: Random Sampling		Random Sampling	257
Lesson 12: Bias in Survey Sampling		Bias, Sampling methods	259
Lesson 13: The Confidence Game		Confidence intervals	262
Lesson 14: How Confident Are You?		Confidence intervals, margin of error	265
Lab 3D: Are You Sure about That?		Bootstrapping	268
Practicum: Let's Build a Survey!		Non-biased survey design	271
Section 4: What's the Trigger?			272
Lesson 15 Ready, Sense, Go!		Sensors, data collection	274
Lesson 16: Does it have a Trigger?		Survey questions, sensor questions	277
Lesson 17: Creating Our Own Participatory Sensing Campaign		Participatory Sensing campaign creation	280
Lesson 18: Evaluating Our Own Participatory Sensing Campaign		Statistical questions, evaluate campaign	284
Lesson 19: Implementing Our Own Participatory Sensing Campaign	Class Campaign—data	Mock-implement campaign, campaign creation, data collection	286
Section 5: Webpages			288
Lesson 20: Online Data-ing	Class Campaign—data	Data on the internet	290
Lab 3E: Scraping web data	Class Campaign—data	Scraping data from the internet	294
Lab 3F: Maps	Class Campaign—data	Making maps with data from the internet	297
Lesson 21: Learning to Love XML	Class Campaign—data	Data storage, XML	299
Lesson 22: Changing Orientation	Class Campaign—data	Converting XML files	301
Practicum: What Does Our Campaign Data Say?	Class Campaign—data	Statistical questions, visualizations, numerical summaries	303
End of Unit Project: TB or Not TB	Class Campaign	Simulation using experiment data	304

Table of Contents (continued)

Unit 4	Campaign	Topics	Page
Daily Overview			307
Essential Concepts			309
Section 1: Predictions and Models			311
Lesson 1: Water Usage		Data cycle, official datasets	313
Lesson 2: Exploring Water Usage		Exploratory data analysis, campaign creation	317
Lesson 3: Evaluating and Implementing a Water Campaign	Water Campaign—data	Statistical questions, evaluate & mock implement campaign	319
Lesson 4: Refining the Water Campaign	Water Campaign—data	Revise and edit campaign, data collection	322
Lesson 5: Statistical Predictions using One Variable	Water Campaign—data	One-variable predictions using a rule	324
Lesson 6: Statistical Predictions by Applying the Rule	Water Campaign—data	Predictions applying mean square deviation, mean absolute error	327
Lesson 7: Statistical Predictions using Two Variables	Water Campaign—data	Two-variable statistical predictions, scatterplots	330
LAB 4A: If the Line Fits…	Water Campaign—data	Estimate line of best fit	332
LAB 4B: What's the score?	Water Campaign—data	Comparing predictions to real data	334
Lesson 8: What's the Trend?	Water Campaign—data	Trend, associations, linear model	336
Lesson 9: Spaghetti Line	Water Campaign—data	Estimate line of best fit, single linear regression	340
LAB 4C: Cross-Validation	Water Campaign—data	Use training and testing data for predictions	343
Lesson 10: Predicting Values	Water Campaign—data	Predictions based on linear models	345
Lesson 11: How Strong Is It?	Water Campaign—data	Correlation coefficient, strength of trend	348
LAB 4D: Interpreting Correlations	Water Campaign—data	Use correlation coefficient to determine best model	350
Section 2: Piecing It Together			353
Lesson 12: More Variables to Make Better Predictions	Water Campaign—data	Multiple linear regression	355
Lesson 13: Combination of Variables	Water Campaign—data	Multiple linear regression	358
LAB 4E: This Model is Big Enough for All of Us	Water Campaign—data	Multiple linear regression	361
Practicum: Predictions	Water Campaign—data	Linear regression	362
Lesson 14: Improving your Model	Water Campaign—data	Non-linear regression	363
LAB 4F: Some models have curves	Water Campaign—data	Non-linear regression	365
Section 3: The Growth of Landfills			367
Lesson 15: The Growth of Landfills	Water Campaign—data	Modeling to answer real-world problems	369
Lesson 16: Exploring Trash via the Dashboard	Water Campaign—data	Analyze data to improve models	373
Lesson 17: Exploring Trash via RStudio	Water Campaign—data	Analyze data to improve models	374
Prepare Team Presentations	Water Campaign—data	Modeling with statistics	-
Present Team Recommendations	Water Campaign—data	Modeling with statistics	-
Section 4: Decisions, Decisions!			375
Lesson 18: Grow Your Own Classification Tree	Water Campaign—data	Multiple predictors, classifying into groups, decision trees	377
Lesson 19: Data Scientists or Doctors?	Water Campaign—data	Decision trees based on training and testing data	382
LAB 4G: Growing Trees	Water Campaign—data	Decision trees to classify observations	385
Section 5: Ties That Bind			387
Lesson 20: Where Do I Belong?	Water Campaign—data	Clustering, k-means	389
LAB 4H: Finding Clusters	Water Campaign—data	Clustering, k-means	394

Lesson 21: Our Class Network	Water Campaign—data	Clustering, networks	396
End of Unit 3 and 4 Design Project and Oral Presentation: Water Usage	Water Campaign	Synthesis of above	400