Learning Plan Examples

Analytical Methods for Global Policy Studies (PA397G) https://amgps.jima.me/

Focus: Fundamentals, Data Cleaning, and Statistical Thinking in Python

RESOURCES/MATERIALS

DataCamp courses

LEARNING PLAN OBJECTIVE

My main objective is to learn the fundamentals of Python through DataCamp, as I will use Python for the replication project. I will also learn about data cleaning and statistical thinking in Python.

LEARNING PLAN TIMELINE AND ACTION STEPS

Week 1: Sept 19 - Sept 25			
DataCamp Course Work	Documenting Progress/Assessing Performance		
Introduction to Python (4 hrs)	• I will use a running <u>Google Doc</u> to type up my Python notes,		
Python Basics	which I will use for reference and to demonstrate progress		
Python Lists	 Upload screenshots of 4 completed exercises from Data 		
 Functions and Packages 	Camp to <u>Google Drive Folder</u>		
• NumPy			
Week 2: Sept 26 - Oct 2			
DataCamp Course Work	Documenting Progress/Assessing Performance		
Intermediate Python (4 hrs)	Update <u>Google Doc</u> with week's notes		
	Upload screenshots of 4 completed exercises from Data		
Dictionaries & Pandas	Camp to <u>Google Drive Folder</u>		
 Logic, Control Flow, and Filtering 			
Loops			
Week 3: Oct 3 - Oct 9	1		
DataCamp Course Work	Documenting Progress/Assessing Performance		
Dither Date Grieger Teel Day, Dart 4 (2 has)			
Python Data Science Tool Box - Part 1 (3 hrs)	Update <u>Google Doc</u> with week's notes		
• writing your own functions	Upload screenshots of 3 completed exercises from Data Complete Concerns to Concer		
 Default arguments, variable-length arguments and scene 	Camp to <u>Google Drive Folder</u>		
Lambda functions and error bandling	 The case study looks very helpful for practicing skills thave learned thus far Lam including this task as "ontional" 		
	because I should still be able to successfully complete the		
Intermediate Bythen (OPTIONAL if time permits)	final skills assessment on DataCamp without it.		
Case Studiu Hacker Statistics			
Case Study: Hacker Statistics			
Dete Course Course Wards			
DataCamp Course work	Documenting Progress/Assessing Performance		
Buthon Data Science Tool Box Dart 2 (4 brs)	 Undate Coople Dec with week's notes 		
Python Data Science Tool Box - Part 2 (4 hrs)	Update <u>Google Doc</u> with week's notes		
 Python Data Science Tool Box - Part 2 (4 hrs) Using iterators in PythonLand 	 Update <u>Google Doc</u> with week's notes Upload screenshots of 2 completed exercises from Data 		

Week 5: Oct 17 - Oct 23			
DataCamp Course Work	Documenting Progress/Assessing Performance		
Python Fundamentals Final Skills Assessment	 Complete Final Skills Assessment and upload screenshot to Google Drive Folder 		
Introduction to Importing Data in Python (3 hrs)	Update Google Doc with week's notes		
Introduction and flat files	Upload screenshots of 3 completed exercises from Data		
 Importing data from other file types 	Camp to Google Drive Folder		
 Working with relational databases in 			
Python			
Week 6: Oct 24 - Oct 30			
DataCamp Course Work	Documenting Progress/Assessing Performance		
Cleaning Data in Python (4 hrs)	• Update <u>Google Doc</u> with week's notes		
Common data problems	Upload screenshots of 4 completed exercises from Data		
 Text and categorical data problems 	Camp to <u>Google Drive Folder</u>		
 Advanced data problems 			
Record linkage			
Week 7: Oct 31 - Nov 6			
DataCamp Course Work Documenting	Progress/Assessing Performance		
Statistical Thinking in Python Part 1 (3 hrs) •	Update <u>Google Doc</u> with week's notes		
Graphical Exploratory Data Analysis	Upload screenshots of 4		
completed exercises from Data			
Quantitative Exploratory Data Analysis	Camp to <u>Google Drive Folder</u>		
Thinking Probabilistically- Discrete			
Variables			
Ininking Probabilistically- Continuous Variables			
Week 8: Nov 7 - Nov 13			
DataCamp Course Work	Documenting Progress/Assessing Performance		
Statistical Thinking in Python Part 2 (3 hrs)	Update Google Doc with week's notes		
Parameter estimation by optimization	Upload screenshots of 4 completed exercises from Data		
Bootstrap confidence intervals	Camp to <u>Google Drive Folder</u>		
Introduction to hypothesis testing	• Submit final assessment of my own performance using the		
Hypothesis test examples	above evaluation plan to Canvas.		
Putting it all together: a case study			

Summary: The Python Fundamentals Final Skills Assessment will be integral in allowing me to assess my performance on mastering the fundamentals of Python (which include both basic and intermediate skills). Because I plan on finishing the entire "Python Fundamentals" course series on Data Camp, I feel confident that I will receive a passing grade on the assessment. Furthermore, I plan on developing my data cleaning and hypothesis testing skills through four additional courses. I chose these courses due to their relevance to the course and application to the replication project.

Semester Objective and Professional Reasoning: multivariate regression analysis, but along the way manipulate data from mixed methods studies as well as be comfortable with data visualizations (applicable to public policy, e.g. 2-dimensional density plot with viridis package multi-colored/ scatter density plot)

Method to Achieve: DataCamp Courses and Assessments for the major outcomes

Proof of Success: First, a screenshot of each course completed and exercises completed. Second, the ability to contribute to the replication project and my research in other classes by utilizing skills in R.

	Objective During Week	Assessment
9/21	Introduction to R	Complete course in Data Camp
		(4hrs) and explain action items
		to partner
9/28	Intermediate R	Complete course in Data Camp
		(6hrs) and explain action items
		to partner
10/5	Intro to Statistics with R	Statistics Fundamentals in R
		Assessment After Data Camp
		Course (4hrs) and explain
		action items to partner
10/12	Intro to Regression with R	Data Camp Course (4hrs) and
		explain action items to partner
10/19	Intermediate Regression with R	Data Camp Course (4hrs)
		explain action items to partner
10/26	Intro to Data Visualization with ggplot2	Data Camp Course (4hrs) and
		explain action items to partner
11/2	Data Visualization in R (comprehensive)	Data Camp Course (4hrs) and
		explain action items to partner
11/9	Week to work on replication and improve weak	Data Visualization with R
	skill areas	Assessment and explain action
		items to partner

Action Items achievable in each course

Introduction to R

- 1. Learn how to use the console as a calculator and how to assign variables. Also know the basic data types in R
- 2. Create vectors in R, name them, select elements from them, compare different vectors
- 3. Create matrices and understand how to do basic computations with them
- 4. Create, subset, and compare factors (categorical data) in R
- 5. Create data frames, highlight specific variables and order them in meaningful matters
 6. Create, subset, and name lists in R

Intermediate R

- 1. Build conditional statements by learning about relational objects in R and how to create "and" / "or" in logical operators
- 2. Learn about 'if loops' and 'for loops' and how to use them appropriately
- 3. Create functions in R and practice making my own
- Learn how and when to use 'lapply' function over a vector/list
- Learn functions to structure and manipulate data including 'time' and 'date'

Statistics Fundamentals in R

- 1. Identifying common statistical distributions and simulating appropriate data
- 2. Understanding and implementing hypothesis tests
- 3. Understanding and implementing sampling techniques and their use in analysis
- 4. Using graphics to learn about the distribution of data and validate assumptions

Introduction to Regression in R

- 1. Fit simple linear regression models with numeric and categorical explanatory variables, and how to describe the relationship between the response and explanatory variables using model coefficients
- 2. Learn regression to the mean and how to transform variables
- 3. Learn to quantify the fit of model and diagnose issues using visualizations as well as observations influence and leverage on the model
- 4. Learn simple logic regression and assessment using confusion metrics Intermediate Regression in R
- 1. Learn parallel slopes regression with one numeric and one categorical explanatory variable
- 2. Deal with Simpson's paradox with multiple explanatory variables
- 3. Implement linear regression algorithm for two explanatory variables
- 4. Understand logistic distribution and be able to execute a logistic regression algorithm Intro to Data Visualizations in R with ggplot2

- 1. Understand the 7 grammatical layers and aesthetic mapping
- 2. Convert categorical and continuous variables into visual scales
- 3. Know the geometrics in R to create bar charts, scatter plots, and line plots

Data Visualization with R

- 1. Creating visualizations suited to numeric, categorical, or time series data types
- 2. Creating statistical visualizations
- 3. Creating plots on data-aware grids
- 4. Customization of plots to optimize communication of insights

Interactive Data Visualization with Plotly Bonus over Fall and Spring Semester break

https://www.sharpsightlabs.com/blog/density-plot-in-r/

Methods Learning Plan Proposal

My main goals for the semester:

- Expand my comfort zone skills: Gain skills in excel.
- Expand out of comfort zone skills: Gain beginner skills in Python
- Challenging skills I may not finish: Intermediate skills in Python Methods for reaching

my goals:

- Datacamp Modules
 - \circ Spreadsheet Fundamentals
 - \circ Python Fundamentals

Week	Work for the week	Way to track progress
9/14-9/21	Datacamp: Data Analysis in Spreadsheets (3 hours)	Screenshot of finished modules
9/21-9/28	DataCamp: Intermediate Spreadsheets (4 hours)	Screenshot of finished modules
9/28-10/5	DataCamp: Pivot Tables in Spreadsheets (4 hours)	Screenshot of finished modules
10/5-10/12	DataCamp: Data Visualization in Spreadsheets (4 hours)	Data Camp: Statement of Accomplishment Earned
10/12-10/19	DataCamp: Introduction to Python (4 hours)	Screenshot of finished modules
10/19-10/26	DataCamp: Intermediate Python (4 hours)	Screenshot of finished modules
10/26-11/2	DataCamp: Python Data Science Toolbox Part 1(4 hours)	Screenshot of finished modules
11/2-11/12	DataCamp: Python Data Science Toolbox Part 2(4 hours)	Python Programming Skill Assessment (screenshot results)
		Statement of Accomplishment Earned

Assessing my performance:

By November 12, 2021, I should have two Statements of Accomplishments from DataCamp along with screenshots from each week documenting my progress throughout the semester.

Objectives

My personal objectives for this learning plan throughout the course of the Analytical Methods class are the following:

- Be able to understand and master the basic applications of Microsoft Excel. These include the building blocks for data analysis, including, but not limited to, the basic functions of rows, columns, graphs, and formulas.
- Be able to summarize data based on measures of central tendency and dispersion.
- Be able to calculate and interpret z-scores of a distribution through Excel.
- Develop proficiency in statistical analysis and application of statistical terms in Excel.

These objectives are complementary to my career goals, which include working in monitoring and evaluation within the field of International Development. Furthermore, my background in organizing and program implementation during my undergraduate career allowed me to identify the gaps in my learning and skills, most of which pointed to a lack of familiarity with softwares like Excel. Not only do the aforementioned objectives tackle these gaps, but their design will allow me to use quantitative analysis throughout the rest of my career. Furthermore, for the purposes of the replication project, which requires exporting data from sources like the Heritage Foundation, the Organization for Economic Cooperation and Development, and the World Bank, Excel will prove to be a useful tool in conducting regression analysis (Quazi, Ballentine, Bindu, and Blyden 2018, 286-287).

Week of September 27, 2021

DataCamp - "Introduction to Spreadsheets" |Major Concepts: Basic working knowledge of Excel - rudimentary stages

• Throughout the first week of the learning plan, I will gain a thorough understanding of the most basic functions in Spreadsheets, including, not limited to, rows and columns, formulas, and exponents.

Weeks of October 4, 2021 and October 11, 2021

DataCamp - "Data Analysis in Excel" | Major Concepts: Introduction to Excel

• Throughout this course, I will learn how to use math functions, as well as Excel functions such as VLOOKUP, AVERAGEIF(S), and keyboard shortcuts. This will serve as a preliminary introduction to manipulating data, which will take place starting October 18. The contents of the course are "Exploring Data, Preparing Data," and "Analyzing Data" (DataCamp, n.d.).

Weeks of October 18, 2021 and October 25, 2021

DataCamp - "Introduction to Statistics in Spreadsheets" | Major Concepts: Statistical Analysis and Visualization • Throughout this course, I will learn to conduct hypothesis testing, visualize z-scores, apply statistical analysis in regards to measures of central tendency, correlations, and learn to manipulate graphics such as Histograms. The course contents include "Getting To Know Your Data, Statistical Data Visualization," and "Statistical Hypothesis Testing" (DataCamp, n.d.).

Weeks of October 25, 2021 and November 1, 2021

Datacamp - "Pivot Tables in Spreadsheets" | Major Concepts: Organizing large data sets

• This course will give me a thorough understanding of pivot tables, including, but not limited to, organizing large data sets and manipulating minimum and maximum values, as well as measures of central tendency. The course contents include "Pivot Table Introduction in Sheets, How to manipulate pivot tables and data," and "Advanced analysis in pivot tables" (DataCamp, n.d.).

Weeks of November 8

DataCamp - "Intermediate Spreadsheets" | Major Concepts: Applying transformations of functions in Spreadsheets

Throughout the last week of this learning experience, I will learn how to conduct transformations of functions, including logarithmic, exponential, and square root transformations. This will aid in the process of regression analysis in the replication plan, which includes logarithmic transformations (Quazi, Ballentine, Bindu, and Blyden 2018, 286). Because this module includes three sections of course contents, given the weeklong timeframe to complete the learning plan, I will complete the second module of the course. The course contents include "Working with Numbers" (DataCamp, n.d.).

Assessing Completion

My assessment of my performance during the learning plan will be based on the following criteria:

- Provide documentation/proof of completion of the assigned modules (screen shots, PDFs)
- 1 screenshot/PDF for each completed module
- Provide a short paragraph or reflection about the overall satisfaction of the learning plan
- Provide proof of data manipulation through Excel by using the datasets used from the replication plan
- To measure performance, I will:
- Use VLOOKUP and AVERAGEIF(S) on the replication dataset
- Create a histogram with replication dataset
- Show manipulation of a pivot table
- Create a transformation of a function

Works Cited

- Quazi, R. M., Ballentine, W. E., Bindu, F., & Blyden, L. (2019). Multilateral foreign aid, bilateral foreign aid, and foreign direct investment in Latin America. *International Journal of Economics and Financial Issues*, *9*(2), 284-290.
- DataCamp. n.d. "Data Analysis in Excel." Accessed September 14, 2021. https://learn.datacamp.com/courses/data-analysis-in-excel
- DataCamp. n.d. "Intermediate Spreadsheets." Accessed September 14, 2021. <u>https://learn.datacamp.com/courses/intermediate-spreadsheets</u>
- DataCamp. n.d. "Introduction to Spreadsheets." Accessed September 14, 2021. https://learn.datacamp.com/courses/introduction-to-spreadsheets
- DataCamp. n.d. "Introduction to Statistics in Spreadsheets." Accessed September 14, 2021. https://learn.datacamp.com/courses/introduction-to-statistics-in-spreadsheets
- DataCamp. n.d. "Pivot Tables in Spreadsheets." Accessed September 14, 2021. https://learn.datacamp.com/courses/pivot-tables-in-spreadsheets

Learning Goals:

To become proficient at cleaning, manipulating, and designing data for human visualization and consumption through the understanding of foundational methods and Python computational tools.

Cleaning Data:

Sort through data for relevant information

- After development of the replication projects outline, I will collect the data and work with the team for two weeks to clean it.
- Data will be cleaned with succinct documentation. A copy of the original data set will be saved for reference.

Organize concept appropriately

The new tidy dataset will be rearranged to be understood by Python instead of humans. Data will not have missing values that will negatively affect the study.

Designing Data for Visualization:

Make data logical, engaging, and readable for humans

Adjust outline to include appropriate data visualizations for chosen replication project.

Create cohesive and attractive bar graphs, histograms, heat maps, area charts, and highlight table Create common color palate and font choices to go between all group members

Understanding Methods:

Learn foundations for analyzing through weekly readings

- Development of an editable replications outline citing key weekly readings for future reference Once the replication project research paper has been picked, I will work with members of the team to build a flexible outline for the methods project including not only our methodologies and who is working on what at what time, but also what data visualizations we hope to develop during the process.
 - This is an ongoing project and will be expected to be referenced and updated weekly after article selection

Python:

Readable Functions | Comprehensions | Modules | Scopes and Namespaces | File Handling | Object-Oriented Programming (OOP) | Become familiar with the Pandas Library | Iterator, Generator, Decorators | Functions | Legible Documentation | Succinct Code | Merging Datasets | R Integration (rpy2)

Read Python for Data Analysis (PDA), 2nd Edition by Wes McKinney an hour a week.

Timeline:

Key:

DataCamp: DC | *Python for Data Analysis*: PDA | Data Visualization: DV | Replication Project: RP

Week	Class Homework	Self-Learning/RP Actions	Coding Focus
9/14	Complete readings, draft learning plan	2-hour RP, 1 hour group meetup, attend Data and Donuts on Managing Research Data, office hour on Friday	2 Hour DC, 1 Hour PDA reading
9/21	Complete readings, finish learning plan proposal	2-hour RP, complete readings, confirm RP research	2 Hour DC, 1 Hour PDA reading
9/28	Complete readings: Probability and sampling, check chosen research for proper sampling methods	2-hour RP (collecting data, developing outline), 1 hour group meetup	2 Hour DC, 1 Hour PDA reading, submit first assessment screenshot
10/05	Complete readings: Inferential statistics & Hypothesis testing	2-hour RP (editing ouotline, cleaning data)	2 Hour DC, 1 Hour PDA reading
10/12	Complete readings: Inferential statistics II - Population and testing difference	2-hour RP (coding for math), 1 hour group meetup	2 Hour DC, Finish PDA
10/19	Complete readings: Bivariate linear regression	2-hour RP (coding for math, coding for visualization)	2 Hour DC
10/26	Complete readings: Multivariate regression,	2-hour RP (coding for visualization), 1 hour group meetup	2 Hour DC, begin DV if ready
11/02	Complete 11/02 readings and 11/09 readings: Introduction to computational social science methods, Introduction to computational social science methods	3-hour RP (finish coding, manipulating data), write draft of replication project	4 Hour DC, submit second assessment screenshot, cont., work on DV
11/09	I am getting married this week and will be unable to attend class. I will be turning in homework 11/02.	N/A	N/A
11/16	Complete readings: Best practices of fieldwork for qualitative researchers	3-hour RP, 1.5-hour group meetup, edit draft of replication report	2 Hour DC, data visualizations, cont. work on DV
11/23	Attend project consultation	3-hour RP, 1.5-hour group meetup, edit draft of RP, practice RP	2 Hour DC, cont. DV, DC project
11/30	RP presentation	Present RP, Verification report of other group's replication study	2 Hour DC, turn in finished DC project

Performance Assessment:

Final Python code utilizes at least two tidy data sets resulting in readable, well organized data displays for human understanding. I will make attractive and coherent data visualizations for the replication project. I will complete 2 assessments and one project for the DataCamp's Python Data Scientist career track and submit screenshots.

R Learning Plan Proposal

Learning Goal:

- My goal in this learning plan is to develop a baseline understanding of R and subsequent topics related to the use of R in data analysis. In order to do this, I will be utilizing courses on DataCamp.

Timeline and Tasks:

- *Week 1: 9/27 10/3* Introduction to R (4hrs)
 - □ Intro to basics
 - I Vectors
 - I Matrices
 - I Factors
 - Data frames
 - Lists
- *Week 2: 10/4 10/10* \circ Introduction to
 - Importing Data in R (3hrs)
 - □ Importing data from flat files with utils
 - Readr & data.table
 - Importing Excel data
 - IReproducible Excel work with XLConnect
- *Week 3: 10/11 10/17* \circ Introduction to
 - Regression in R (4hrs)
 - □ Simple linear regression
 - D Predictions and model objects
 - □ Assessing model fit
 - □ Simple logistic regression
- *Week 4: 10/18 10/24* \circ Introduction to Data

Visualization with ggplot2 (4hrs)

- □ Introduction
- Aesthetics
- Geometries
- I Themes
- *Week 5: 10/25 10/31* \circ Cleaning Data in R
 - (4hrs)
- Common data problems
- □ Categorical and text data □ Advanced data problems
- Record linkage
- *Week 6: 11/1 − 11/7* Introduction to the

Tidyverse (4hrs)

- Data wrangling
- Data visualization

- □ Grouping and summarizing
- Image: Types of visualizations**Assessment:**
- The success of my learning plan will be determined by weekly screenshots of my completion status in each course section.

Customized Learning Plan

8 weeks beginning September 15 and Ending November 3

Overview:

Upon realizing my Excel skills are lacking more than I realized, I will spend most of my time learning data analysis through Excel, especially because it is most widely used in a variety of jobs. Then, because I don't have enough knowledge about other analysis platforms to know which I might prefer to use, I will do one week each of R, Python, and Stata. All except Stata will be completed through DataCamp courses. Stata will be completed via a webinar posted on YouTube. Progress will be assessed based on completion of DataCamp modules and Webinar, for which I will have to show completion though notes and/or screenshots of work in Stata.

Objectives:

1. Gain essential intermediate knowledge of Excel functions, visualization, and data analysis.

2. Gain basic knowledge of R, Python, and Stata to see if I find one more useful or preferable.

	Excel Fundamentals	Tracking Progress
Week 1 Friday 9/15	 Introduction to Excel (3 Hours) Complete Chapter 1: Getting Started with Excel Complete Chapter 2: Managing data and applying aggregate functions 	Submit screenshots for proof of completed chapters.
Week 2 Thursday 9/21	 Data Preparation in Excel (4 Hours) Complete Chapter 1: Starting Data Preparations in Excel Complete Chapter 3: Conditional Formulas Complete Chapter 4: Lookups and Data Transformation 	Submit screenshots for proof of completed chapters.
Week 3 Friday 9/29	Return to skipped chapters in Introductionto Excel and Data Preparation in Exceland Start Data Visualization in Excel (4Hours)-Complete Chapter 3: Other functions and visualizing date-Complete Chapter 2: Functions of Data Preparation-Complete Chapter 1: Building Basic Charts	Submit screenshots for proof of the completed chapters.

Week 4 Friday 10/6	 Finish Data Visualization in Excel (4 Hours) Complete Chapter 2: Advancing to more complex charts Complete Chapter 3: Data visualization best practices Complete Chapter 4: Visualizing disaggregated data with PivotCharts 	Submit screenshots for proof of completed chapters.
Week 5 Friday 10/13	 Data Analysis in Excel (3 Hours) Complete Chapter 1: Exploring data with PivotTables Complete Chapter 2: Intermediate Logical Functions 	Submit screenshots for proof of completed chapters.
	R Programming	
Week 6 Friday 10/20	 Introduction to R (4 Hours) Complete Chapter 1: Intro to Basics Complete Chapter 2: Vectors Complete Chapter 3: Matrices Complete Chapter 4: Factors 	Submit screenshots for proof of completed chapters.
	Python Fundamentals	
Week 7 Friday 10/27	 Introduction to Python (4 Hours) Complete Chapter 1: Python Basics Complete Chapter 2: Python Lists Complete Chapter 3: Functions and Packages Complete Chapter 3: NumPy 	Submit screenshots for proof of completed chapters.
	Stata for Data Analysis	
Week 8 Friday 11/ 3	 UCLA Intro to Stata Webinar (3 Hours) Navigating Stata Data Import Exploring Data Data Visualization Data Management Basic Statistical Analysis 	Submit notes or summary of what was learned.



Analytical Methods Learning Plan Proposal

Resources: DataCamp and LBJ Statistical Workshops

Goals:

- 1. Gain beginner skills in R.
- 2. Complete first half of the "Data Analyst with R" track.
 - a. DataCamp modules in green text are a part of the "Data Analyst with R" track.
- 3. Begin intermediate skills in R (with option to work on Regression in R).

Relevance:

- 1. Important in research to know how to efficiently work with data using R.
- 2. Prepares me for using large, real-world data sets in the workplace.
- 3. Provides foundational knowledge for further programming and statistical analysis.

Assessment/Progress:

- 1. Weekly screenshots as record of completion.
- 2. Complete "Data Manipulation with R" assessment on DataCamp.

Week 1: 9/24- 9/30

DataCamp:

- 1. Introduction to R (4 hours)
 - a. Intro to basics
 - b. Vectors
 - c. Matrices
 - d. Factors
 - e. Data frames
 - f. Lists

Week 2: 9/1-10/7

DataCamp:

- 1. Introduction to Importing Data in R (3 hours)
 - a. Importing data from flat files with utils
 - b. Readr and data.table
 - c. Importing Excel data
 - d. Reproducible Excel work with XLConnect

LBJ Statistical Software Workshop:

2. R Workshop #3 (1.5 hours)

Week 3: 10/8-10/14

DataCamp:

- 1. Introduction to the Tidyverse (4 hours)
 - a. Data wrangling
 - b. Data visualization
 - c. Grouping and summarizing

Week 4: 10/15-10/21

DataCamp:

- 1. Data Manipulation with dplyr (4 hours)
 - a. Transforming data with dplyr
 - b. Aggregating data
 - c. Selecting and transforming data
 - d. Case study: The babynames Dataset

Progress Tracking:

Progress Tracking:

"Progress Tracker"

GoogleDoc.

Compile screenshots of

completed modules into

Compile screenshots of completed modules into "Progress Tracker" GoogleDoc.

"Progress Tracker" GoogleDoc.

Progress Tracking:

Compile screenshots of

completed modules into

Compile screenshots of completed modules into "Progress Tracker" GoogleDoc.

Progress Tracking:

Week 5: 10/22-10/28			
DataCamp: 1. Joining Data with dplyr (4 hours) a. Joining tables b. Left and right joins c. Full, Semi, and Anti Joins d. Case Study: Joins on Stack Overflow Data	Progress Tracking: Compile screenshots of completed modules into "Progress Tracker" GoogleDoc.		
Week 6: 10/29-11/4			
DataCamp:1. Introduction to Statistics in R (4 hours)a. Summary Statisticsb. Random Numbers and Probabilityc. More Distributions and the CLTd. Correlation and Experimental Design	Progress Tracking: Compile screenshots of completed modules into "Progress Tracker" GoogleDoc.		
Week 7: 11/5-11/11			
DataCamp: 1. Introduction to Data Visualization with ggplot2 (4 hours) a. Intro b. Aesthetics c. Geometries d. Themes	Progress Tracking: Compile screenshots of completed modules into "Progress Tracker" GoogleDoc. Assessment: Data Manipulation with R		
Week 8: 11/12-11/18			
DataCamp: 1. Intermediate R (6 hours) a. Conditionals and Control Flow b. Loops c. Functions d. The apply family e. Utilities	Progress Tracking: Compile screenshots of completed modules into "Progress Tracker" GoogleDoc.		
OPTIONAL Week 9: 11/19-11/25			
DataCamp:1. Introduction to Regression in R (4 hours)a. Simple linear regressionb. Predictions and model objectsc. Assessing model fitd. Simple logistic regression	Progress Tracking: Compile screenshots of completed modules into "Progress Tracker" GoogleDoc.		

Learning Plan Proposal

Goals:

- Broaden and solidify my understanding of Excel for data analysis and visualization
- Gain exposure to R for data manipulation and visualization

Resources:

- DataCamp, LBJ Statistical Workshops
- <u>Google Drive</u> (track progress, keep notes)

Week	Topic	Assignments	Assessment
9/4	Excel: Data PreparationR: Introduction	 DataCamp: Starting Data Preparation, Conditional Formulas (1.5 hr) LBJ R Workshop #1 (1.5hr) 	Screenshot of progress
9/11	 Excel: Data Preparation (cont.) Excel: Data Analysis 	 DataCamp: Functions for Data Preparation, Lookups and Data Transformation (1.5 hr) DataCamp: Exploring data with PivotTables, Intermediate Logical Functions (1.5 hr) 	 Statement of Accomplishment Screenshot of progress
9/18	 Excel: Data Analysis (cont.) R: Vectors, data frames, and tidyverse 	 DataCamp: What-if analysis, Forecasting (1.5 hr) LBJ R Workshop #2 (1.5 hr) 	• Statement of Accomplishment
9/25	Excel: Data Visualization	 DataCamp: Data Visualization in Excel (3 hr) 	• Statement of Accomplishment
10/2	R: Programming Introduction	 LBJ R Workshop #3 (1.5 hr) DataCamp: Intro to basics, Vectors, Matrices, Factors (2 hr) 	 Screenshot of progress
10/9	R: Programming Introduction and Intermediate	 DataCamp: Data frames, Lists (2 hr) DataCamp: Conditionals and Control Flow, Loops (2 hr) 	 Statement of Accomplishment Screenshot of progress
10/16	• R: Programming Intermediate (cont.)	• DataCamp: Functions, The Apply Family, Utilities (4 hr)	• Statement of Accomplishment
10/23	• R: Data Manipulation and Visualization	• DataCamp: Introduction to the Tidyverse (4 hr)	Statement of Accomplishment
10/30	R: Data Manipulation and Visualization II	• DataCamp: Data Manipulation with dplyr (4 hr)	• Statement of Accomplishment

1. Analysis of Background, Career Goals, Learning Needs

From a Data Analysis perspective, I do not have any formal training or background experience with Python but I have completed a Coursera Certificate on Data Analytics which addressed the uses of Spreadsheets, SQL (BigQuery), R, and Tableau. Furthermore, I have worked for the past three years with Data Engineers, so I am aware of some concepts of Python such as the data types, manipulation of strings, iterating over objects using loops, task automation, and defining custom classes and methods. However, since I have not yet had a direct need to *use* these concepts in my role, I do not feel that I have retained them - which I hope that this learning plan will address.

In my current role, I do have a direct need for learning how to automate the cleaning/tidying/processing/upload/visualization of data in CSV and JSON form. As for my future career goals, I plan to first achieve a role as a data analyst at a 'Big Tech' company such as Google or Meta. Upon examining the current job descriptions for these types of roles, my learning needs include the following:

- Correlating and attributing user activity based on technical and geopolitical variables
- Tracking users via network infrastructure and malicious code artifacts
- Technical Analysis, including the identification and prioritization of risks based on metadata
- Developing tools to uncover anomalous activity leveraging large data sets
- Scripting for the purpose of automating the tasks involved with security incident escalation

Based on these role requirements, I have defined the following objectives for myself:

1. Learning Goals/Objectives

- Learn how to distinguish between different types of Python data types
- Develop proficiency in parsing text, integers, and dates
- Cultivate understanding of capabilities/limitations of Python libraries such as NumPy, Pandas, and Matplotlib and how they complement each other
- Acquire ability to make API calls using Python
- Demonstrate competency in defining and calling functions in Python

Week (of)	Class Topic	Self-Learning Activities
09/24	Probability and Sampling	DC - Introduction to Python

2. Milestone Schedule

(W6)		(4 hours)
10/01	Inferential Statistics I -	DC - <u>Data Visualization with Matplotlib</u>
(W7)	Hypothesis Testing	(4 hours)
10/08	Inferential Statistics II -	DC - <u>Intermediate Python</u>
(W8)	Testing Difference	(4 hours)
10/15	Inferential Statistics Recap	DC - <u>Importing Data with Python</u>
(W9)	and Applications	(3 hours)
10/22 (W10)	Regression I - Testing relation between two variables	DC - <u>Data Manipulation with Pandas</u> (4 hours)
10/29 (W11)	Regression II - Testing relation between multiple variables	DC - <u>Working with Dates and Times</u> (4 hours)
11/05 (W12)	Regression recap and applications	DC - <u>Regular Expressions in Python</u> (4 hours)
11/12	Consultation Session (No	DC - <u>Web Scraping with Python</u>
(W13)	Class)	(4 hours)

3. Performance Assessment/Evaluation Plan

On a weekly basis, I can perform the following tasks to validate my progress and assess my performance for this assignment:

- Uploading my notes from each learning opportunity to a Google Doc
- Uploading screenshots of lesson/module completion to Google Drive
- Write a paragraph summary of lessons learned for each module, including what concepts are still unclear that I should pursue further *and* how I can imagine I might use concepts I learn in my future career tasks
- Demonstrate an ability to manipulate the contents of CSV, JSON, and potentially XML files using Python for the purpose of follow-on visualization

Learning Plan Proposal Draft

Semester Objective and Professional Reasoning: To gain a basic understanding of the 'R' programming language and how it applies to data analysis as well as mastering Excel. This is the programming language that will most benefit me in my career as a policy analyst or consultant.

Method to Achieve: LBJ, Brilliant and DataCamp Courses

Proof of Success: A screenshot of each course completed, and exercises completed.

	Objective During Week	Assessment
9/28	Intro to Excel Course & R Workshop	LBJ and Data Camp Course
		(4 Hours)
		Submit Screenshots and
		Notes
10/5	Data Preparation in Excel	Data Camp Course (4 Hours)
		Submit Screenshots
10/12	R Workshop & Data Analysis Fundamentals	LBJ Course & Brilliant
		Course (4 Hours)
10/10		Submit Screenshot and Notes
10/19	Introduction to R	Data Camp Course (4 Hours)
		Carbon it Company to at
10/10	Laterne dista D	Submit Screenshot
10/16	Intermediate R	Data Camp Course (6 Hours)
		Submit Scroonshot
11/0	Excel Workshop	L BL Course (2 Hours)
11/2	Excel workshop	LBJ Course (2 Hours)
		Submit Notes
11/9	Introduction to Importing Data in R	Data Camp Course (4 Hours)
11/5		
		Submit Screenshot
11/16	Exploratory Data Analysis in R	Data Camp Course (4 Hours)
	r · · · · · y = · · · · · y = · · · · · y	······································
		Submit Screenshot