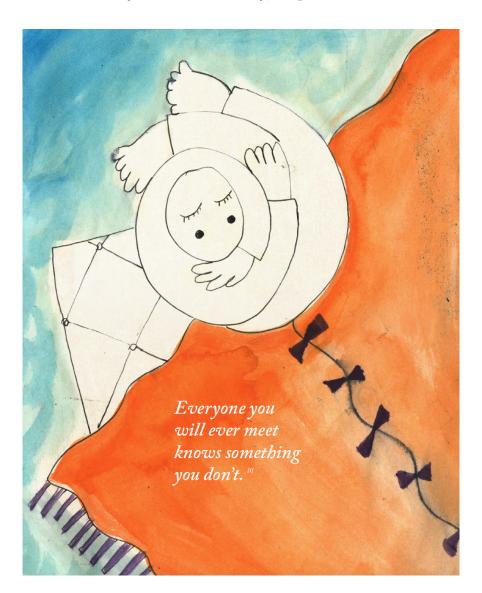## Lab 2. Data Science Basics: Guiding Principles, Component Parts and Analytic Techniques

Due by noon on Friday, September 11th

"Time turns each one of us into another person, and by attempting to communicate with strangers, we help ourselves to communicate with our future selves." Schwab and Claerbout

Objectives of this laboratory:

1. To install the chloropleth package in R and create a map that statistically describes one or more American States county by county

2. To add dataset and create a spatial object in R and dissagregate that political subdivision into its component parts and explore the effect in terms of spatial statistical description

3. To use analytic techniques in order to describe a demographic or economic characteristic of that political subdivision and graphically present that quantitative analysis as a map

Software and Resources you will need or is helpful:

1. The R Project for Statistical Computing otherwise known simply as R. The R framework can be downloaded from https://www.r-project.org/

2. Quick R, an online resource developed by Rob Kabacoff, a statistical consultant and methodologist, that is purposed towards introducing students to the power of R. http://www.statmethods.net/

3. Introduction to R, a pdf presentation created by Phil Spector, Applications Manager and Associate Adjunct Professor at Cal-Berkeley. The pdf is more advanced than the the Quick-R online learning resource, but it covers many of R's basics with regard to functionality and emphasis on data management. https://www.stat.berkeley.edu/˜spector/Rcourse.pdf

4. Packages and datasets as instructed during the laboratory

Grading

This lab will be graded based on 4 sub-steps. In the **first** step, you will specify your own intuitions regarding William and Mary's strengths. In the **second** step, you will identify data which either affirms or refutes those intuitions. In the **third** step, you will visualize that data. In the fourth step, you will briefly contrast your intuitive assumptions to your findings using the data. In each step, you will be graded according to the COLL Curriculum Rubric, found along with your syllabus.

Step 1: Why W&M? (10 points)

Rank the following five reasons you chose to attend William and Mary (from 1 to 5, where 1 was the most important factor, and 5 the least important):

__ It is prestigious.

__ It offered the best value.

__ The Social Scene.

__ The Quality of Education.

__ The Diversity of Student Population.

<u>6</u> I really liked the colors.

Step 2: (Dis)prove yourself. (20 points)

Note your top two most important factors. Using excel (or another, similar program), open the file "WhyWMdata.csv" [1], provided with this lab. This file has a number of statistics about William and Mary. Pick two facts that either support or refute the top two reasons you chose William and Mary. Write these down and briefly describe your reasoning.

Top Reason I chose W&M: _____

Data (Supporting or Refuting; circle one) the reason you chose W&M:

_____

_____

_____

_____

Second Reason: _____

_____

_____

_____

_____

Step 3: Visualize your findings. (50 points)

Now, make a visualization arguing for or against your reasons for accepting the offer to attend William and Mary. You must use all four of the data points highlighted in Step 2 above. To make this visualization, you will need to make a free account at http://www.visme.co/, and follow the below steps:

- Click on "Create new Visme"

- Name your project "W&M Data"

- At the top, click infographic.

- Choose from one of the free infographic templates - just pick what you think will best fit your particular data, as you'll be replacing all the content.

- Wait while visme "Loads the Awesome"

- Double click on the graphics to change the values to reflect your own data.

- Change any text to communicate your argument: either that your reasons for choosing W&M were supported by the data, or not!

Bonus Step: Add in data on another university. (up to 10 bonus points)

Using what you've learned so far, google for another university's "Common Data Set". Add

---

[1]https://www.wm.edu/offices/ir/cds/index.php

one element of the infographic comparing William and Mary to this university to further argue for or against your choice to come to W&M!

Step 4: Reflect on your Findings. (20 points)

Finally, using the infographic you just produced draw a conclusion about how your intuition about William and Mary was - or was not - the same as that reflected in the data. If it was the same, why? Did you look at the data before making your choice? If it was different, why? What led you to believe something other than you observed in the data?

_____

_____

_____

_____

Would this data alone suffice to convince you to attend (or not attend!) W&M? What other data would be useful?

_____

_____

_____

_____