**Data Visualization: Final Exam**

Due Saturday Dec 9, 2017

*100 points*

# Part 1: R (40 points)

# Overview

The homework assignments provided an opportunity for you to practice how to use R for data visualization. This section of the final exam is designed to test your knowledge of R for visualization by asking you to correct and improve existing R scripts for Washington D.C.

# Data

There are two data files for the R script: “dc\_host\_network.csv” and “dc\_listings\_exam.csv”. The first file contains the data for the co-host network and the second file contains the detailed listings for Washington D.C. The host network data has three columns: *host.x, host.y*, and *numReviewers*. The first two columns are the first and the second host respectively. Each record represents an edge, and the edges are weighted by the number of reviewers.

The listings data contains 7,778 records and 93 columns. Each record represents a listings on Airbnb in Washington D.C. The column names are fairly self-explanatory, but any background information would be available on the website: <http://insideairbnb.com/get-the-data.html>.

# Exam

The shell for the R portion of the exam is available on Sakai. The R Markdown file has already been created; you will need to correct the script in order to create the HTML document for submission. There are three parts:

1. *Visual Choices.* There are three charts made with ggplot2 that do not adhere to the principles discussed in *Data Points*, and each chart has three areas for improvement. When you make the changes to the script, you should include a comment to indicate what change you made and why. (More details in the R script.)
2. *Errors.* There are two charts, one made with ggvis and one network graph, with errors that prevent them from running. Use your knowledge of R to troubleshoot this code and make it run. Explain why the original code had these errors, and explain how you fixed it. (More details in the R script.)
3. *Insight.* Using the insights you gathered from the graphs from above, create your own chart with a particular insight. You are welcome to use the package of your choice.

Submission Guidelines

Submit HTML file via Sakai no later than **11:55pm** on **Saturday** **December 9.**

# Grading Rubric

Part 1 and Part 2 will be primarily graded on whether you a) identified the problems and b) solved them and/or improved the code according to the principles. Part 3 will be graded on the following criteria:

* Are your charts accurate to the data? (4 pts)
* Are your visualization choices appropriate for the analysis and data? (2 pts)
* Are your charts attractive, easy-to-read, well-labeled, and easy-to-understand? 2 pts)
* Did you put forth additional effort above and beyond the minimum? For example, did you create more sophisticated charts? (2 pts)

# Part 2: Tableau Dashboard (60 points)

# Overview

This portion will ask you to create two dashboards: a dashboard on the competitive space of Airbnb in your city, and a dashboard tool that could help hosts make a pricing decision.

Use this project as an opportunity to be creative and learn more about the landscape of Airbnb in your assigned city, as well as an opportunity to develop a functional tool for any Airbnb hosts who face a pricing decision over the rate to charge for their room.

# Data

Inside Airbnb collects data from the Airbnb website to understand how Airbnb could affect the residential real estate market. The data are freely available for public use and available online at: <http://insideairbnb.com/get-the-data.html>.

You will be responsible for extracting the data for your assigned city from Inside Airbnb. A description of how to extract that data is below.

**Dashboards**

Each dashboard has different requirements, although there are a couple of common themes. The dashboard titles should change dynamically based on the parameter selection in order to make the dashboard clearer to the reader. Each dashboard should contain a textbox, either with insights about Airbnb in that city or with a comparison of the user’s chosen price to the average price.

1. *Competitive Landscape.* Create a dashboard to discuss the competitive landscape for Airbnb in your assigned city. The dashboard should allow the user to focus on the entire city or just a particular neighborhood and give the user relevant information on the number of available listings. (This is comparable to the background information found on the Inside Airbnb page.) Although there is considerable room for creativity, each dashboard must have three main components:
   1. An interactive map with all the listings
   2. At least two additional charts with information about the available listings
   3. A prominent display with the number of listings for the city and/or neighborhood
2. *Price Tool.* Create a price tool to enable a user to input a price and determine how that price compares to the price offered by similar hosts. This dashboard has three components:
   1. A parameter for the user to input the price, neighborhood, and apartment type
   2. A comparison of the user’s price and the mean prices for those choices
   3. At least 2 charts to show the user their price as compared to the average price for similar host rooms.

**Chart Guidelines**

The charts should be both informative to your analysis question and demonstrate an understanding of the principles of data visualization covered in class. The guidelines for the charts are as follows:

* Each dashboard should have at least **two distinct types of charts**, so each chart must contain a different type of visual that we covered in class.
* Each chart should contain at least **three dimensions of data**, i.e., should describe the multiple variables across different groups.
* Each chart should be appropriately labelled to increase the clarity.
* There is an example Tableau workspace available on Sakai. Feel free to use this as a starting point for your dashboards, but do not copy the dashboards.

# **Submission Guidelines**

You should publish the workbook to Tableau public, and your presentation will be graded on published Tableau workspace. Submit the link to your workspace via Sakai no later than **11:55pm** on **Saturday December 9.**

To publish the workbook to Tableau Public, you will need to create the packaged Tableau workspace for submission, go to Server > Tableau Public > Save to Tableau public.

# Grading Rubric

Each Tableau dashboard will be graded according to the following criteria (30 pts each for 60 pts total):

* Are your charts accurate to the data? (8 pts)
* Are your visualization choices appropriate for the analysis and data? (7 pts)
* Is the title and focus of the dashboard clear? (2 pts)
* Are your charts attractive, easy-to-read, well-labeled, and easy-to-understand? (5 pts)
* Did you put forth additional effort above and beyond the minimum? For example, did you create more sophisticated and/or interesting dashboards? (5 pts)
* How well-supported are your conclusions / recommendations? (3 pts)

**Good luck!**

# Downloading Airbnb Data

The Airbnb data for each city is available online at: <http://insideairbnb.com/get-the-data.html>. Download the detailed listings for your city, a file named *listings.csv.gz* and described as the detailed listings. (Screen shot shown below.)

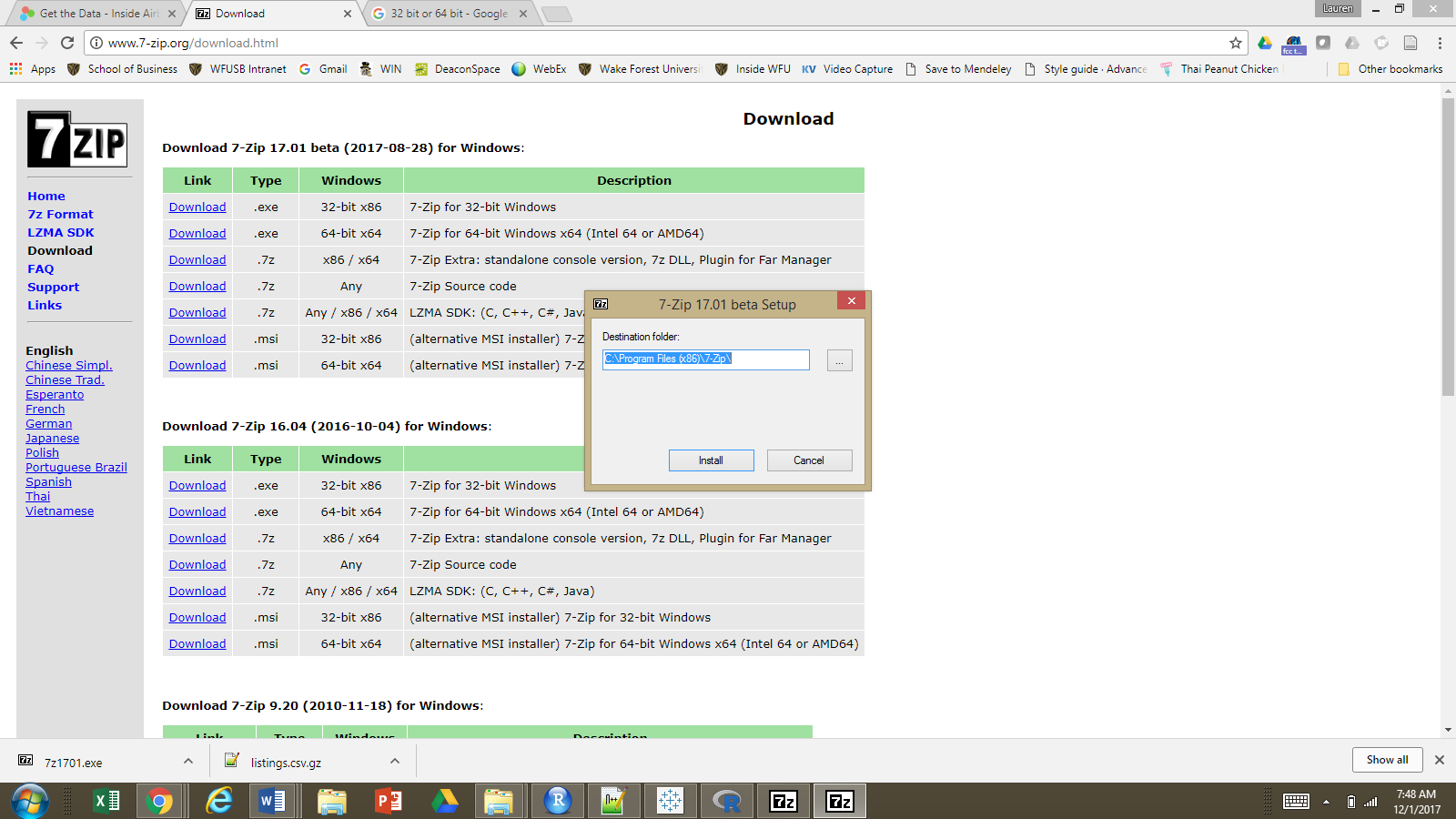


**Extraction**

If you have a Mac, then you can double-click on the .gz file to extract the listings.csv file. If you have a Windows machine, then you will need to download an application to unzip the .gz file. I recommend downloading 7-Zip, available at: <http://www.7-zip.org/download.html>. Click on one of the first two links to download the .exe file for Windows. (Screen shot below.)



After you download the file, follow the prompts to install the program.



Once 7-zip is installed, then you can open the 7-Zip File Manager. Extract the listings\_all.csv.gz file to your folder for the final exam. 7-Zip will create a folder named listings\_all.csv, and the listing data for your city will be in there. (Two screen shots below.)

