Organization of our Course: A Big-Picture Perspective

(A) Be capable of extracting, transforming and loading (ETL) data using multiple platforms (e.g. R & Power BI). **How did we achieve this?**

- **The Bike Sharing Problem:**
  - We used R to <u>extract</u> datasets both from flat files located on your computer and the web. Additionally, we have talked about how to scrape data and use APIs to download data of interest.
  - We used R to <u>transform</u> the dataset. Through the *dplyr* package, we covered the following cleaning steps: (i) filtering data rows; (ii) selecting columns and/or subsets; (iii) adding a calculated column from an existing dataset; and (iv) aggregating data into groups.
  - We then used R to highlight to <u>load</u> the data into a CSV framework so it can be opened using any BI platform.
- **The Airline Data Problem:** We used Power BI to extract, transform and load the data. In that example, we: (i) merged five different data tables; (ii) transformed some of the data columns to a different encoding; and (iii) filtered the data for the purpose of visualizing only the interesting columns. These processes were all performed in Power BI. **In this class, everyone left with their example working** ☺.
- **How did Fadel evaluate our understanding?**
  - Assignments 4-10.
  - **Exam I:** Questions 6-15.
  - **GE Project**
  - **Project:** You needed to do this step in your projects. Obviously, different projects had different needs for the ETL process.

(B) Write basic R scripts to preprocess and clean the data. **While this is part of the ETL, I highlighted this in the syllabus not because R (or any other scripted Language) forces you to think about the logic and process.** <u>See examples above for how we did this and how I evaluated your understanding of it.</u>

(C) Explore the data using visualization approaches that are based on sound human factors (i.e. account for human cognition and perception of data). **How did we achieve this?**

- We learned about Tufte's Graphical Design Excellence, Graphical Integrity and Chartjunk principles.
- We discussed data types and encoding.
- We examined a taxonomy of graphs and learned about the appropriate graph types for different data types. Through that process we examined over 150 examples of "good" and "bad" graphs.
- We emphasized the role of color in understanding your visualized data.
- We spent a **significant portion of classes 15–21 to talk about how Tableau, Power BI, Python (pandas-profiling and pycaret) and R can be used to visualize data according to the aforementioned principles**.
- **How did Fadel evaluate our understanding?**

- Assignments 11-14.
- **Graph and Dashboard Critiques** through: several in-class activities and **Exam II:** Questions 1-16.
- **The effectiveness of utilizing these concepts in your Project and in Exam II: Questions 17-18.**

(D) Understand how data mining and other analytical tools can capitalize on the insights generated from the data visualization process. <u>**How did we achieve this?**</u>

- An overview and motivation for data mining that included several different practical examples from Fadel's funded research projects. In this class, I have also introduced you to the CRISP-DM framework. (see class 24)
- A detailed examination of frequent itemsets & association rules. This included both hand calculations and an example of how to do this in R. (see class 25).
- An introduction to clustering clustering. By hand (your out-of-class activity), using R and using Tableau. Note clustering is extremely useful in visual analytics applications since it is a data-driven way of grouping data. (see class 27).
- <u>**How did Fadel evaluate our understanding?**</u>
  - **Exam III**.

(E) Create interactive dashboards that can be used for business decision making, reporting and/or performance management. <u>**How did we achieve this?**</u>

- We built interactive dashboards using Tableau and PowerBI.
- More importantly, we talked about the importance of reporting on the right metrics. **Recall the paper plane experiment that we have talked about in class**. We also used this experiment to highlight that the right metrics are different to our internal customers (e.g. management vs. engineering).
- <u>**How did Fadel evaluate our understanding?**</u>
  - The project.
  - Class discussion for the Paper Plane Experiment and the associated questions in Exam III.

(F) Be able to apply the skills from this class in your future career. This is why you got to choose your own project ☺.