

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 extract 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 transform 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 load 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?

- Group assignments 01 and 02.
- **Exam I:** Questions 6-15.
- **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. See examples above for how we did this and how I evaluated your understanding of it.**

(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 16–19 and 22 to talk about how Tableau, Power BI and R can be used to visualize data according to the aforementioned principles.** Additionally, in class 26, I introduced you to the use of Python for data visualization (particularly in preparation for predictive modeling and to explain the outcomes from predictive models).

- How did Fadel evaluate our understanding?
 - **Graph and Dashboard Critiques** through: several in-class activities and **Exam II: Questions 1-15.**
 - **The effectiveness of utilizing these concepts in your Project and in Exam II: Question 16.**
- (D) Understand how data mining and other analytical tools can capitalize on the insights generated from the data visualization process. How did we achieve this?
- 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 22)
 - 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 23).
 - A detailed discussion of clustering. We did this by hand, 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 24).
 - How did Fadel evaluate our understanding?
 - **Exam III** - Questions 1, 2, 3, 5, 6 and 7.
- (E) Create interactive dashboards that can be used for business decision making, reporting and/or performance management. How did we achieve this?
- We built interactive dashboard 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).
 - How did Fadel evaluate our understanding?
 - The project.
 - Class discussion for the Paper Plane Experiment and the associated question 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 ☺.