DAV 5400 Spring 2019 Week 9 Assignment (30 Points)

Tidying and Transforming Data

		Los Angeles	Phoenix	San Diego	San Francisco	Seattle
ALASKA	on time	497	221	212	503	1,841
	delayed	62	12	20	102	305
AM WEST	on time	694	4,840	383	320	201
	delayed	117	415	65	129	61

Source: *Numbersense*, Kaiser Fung, McGraw Hill, 2013

The chart above describes arrival delays for two airlines across five destinations. Your task is to:

(1) Create a .CSV file that includes all of the information shown above. You are required to use a "wide" structure similar to how the information appears above, so that you can practice both data tidying and the data wrangling transformations described in Chapter 8 of the "Python for Data Analysis" textbook. Load your .CSV file into you GitHub repository.

(2) Read your .CSV file from your GitHub repository into a Jupyter Notebook, and use your knowledge of combining and reshaping data in Pandas to tidy and transform your data. To get started, think about how you would want the data to appear if it were converted to "long" format, e.g., how would you define a "single observation" for the data shown in the graphic?; How many key values are associated with each data value?; How many columns should your long format structure contain based on the information provided in the graphic shown above?; What would the column headings for the long structure be?; etc. Use your answers to these questions to guide your reshaping/transformational work on the data. **Your reshaping/transformational steps must include converting the above table to a "tidy" long format.** Additional transformational steps (e.g., filling in missing data values, renaming columns, etc.) should be performed as needed to ensure that your data is, in fact, "tidy".

(3) Using your reshaped/transformed data, perform analysis to compare the arrival delays for the two airlines. Some questions you might choose to answer: For each city, which airline had the best on time performance?; Which airline had the best **overall** on time performance?, etc.

(4) Finally, given your "tidy" long format structure, consider what, if any, changes you would make to the visual presentation of the data if you were then asked to transform your "long" data back into a "wide" format: would you mimic the structure of the graphic shown above? If not, how might you transform your "long" data to "wide" format to make its "wide" presentation easier to understand and work with? Provide an example of your recommendation.

Save all of your work for this assignment within **a single Jupyter Notebook** and upload it to your online DAV5400 GitHub directory. Be sure to save your Notebook using the following nomenclature : **first initial_last name_w9_assn**" (e.g., J_Smith_w9_assn).

As a reminder, this assignment is due no later than 11.59pm on Sunday March 24.