Data Set Options

Choose from one of the following data sets or find your own. Additional resources for finding a data set are included at the bottom of this document.

Difficulty Level	Data Set	Overview	Notes
Beginner	**Titanic Data**	Contains demographics and passenger information from a subset of the 2224 passengers and crew on board the Titanic. More information about the data set can be found here.	Create a visualization that shows the demographics or passenger information between those passengers who survived and those who died
Beginner	**Baseball Data**	A data set containing 1,157 baseball players including their handedness (right or left handed), height (in inches), weight (in pounds), batting average, and home runs.	Create a visualization that shows differences among the performance of the baseball players.
Intermediate	**Flights**	The data set which contains information on United State flight delays and performance comes from RITA. You can download the data directly from RITA or as zipped csv files from the Flights link. The files on the Flights link are organized by year and are more compressed than the originals. Additional details about the data can be found at here.	Investigate the performance of flights over time or simply look at data for a given year and create a graphic that showcases your finding(s).
Intermediate	**Loan Data from Prosper** Last updated 03/11/2014 This data dictionary explains the variables in the data set.	This data set contains 113,937 loans with 81 variables on each loan, including loan amount, borrower rate (or interest rate), current loan status, borrower	Ask your own questions about this data set to find interesting trends in the data.

		income, and many others.	
Advanced	**PISA Data** **PISA Data Dictionary** Note: The unzipped PISA Data csv file is 2.75 GB.	PISA is a survey of students' skills and knowledge as they approach the end of compulsory education. It is not a conventional school test. Rather than examining how well students have learned the school curriculum, it looks at how well prepared they are for life beyond school. Around 510,000 students in 65 economies took part in the PISA 2012 assessment of reading, mathematics and science representing about 28 million 15-year-olds globally. Of those economies, 44 took part in an assessment of creative problem solving and 18 in an assessment of financial literacy. The data and topics of investigation come from the PISA Data Visualization Competition. For inspiration and examples, see the winners and submissions here.	Consider creating a graphic that explores one of the following topics. The importance of school factors in explaining academic performance. Differences in achievement based on gender, location, or student attitudes. Differences in achievement based on teacher practices and attitudes. Inequalities in academic achievement.
Varies Depends on your experience working with data.	**Find your own data set!**	**Remember that finding and cleaning your own data set could take significant time and effort!** See the checklist below if you want to choose your own data set.	Pose your own question and find data to answer it. Alternatively, find a data set and ask questions about it until you find something interesting you want to share.

If you're finding your own data set…

The data set that you eventually submit should:

- ❏ be in a tidy format[1] (you may need to clean and reshape the data)
- ❏ be in a commonly used format of loading data with dimple.js or d3.js such as .csv, .tsv, .txt, .json, .xml, or .html

Here are a few resources to find a data set:

- http://www.pewglobal.org/category/datasets/
- http://databank.worldbank.org/data/home.aspx
- http://www.data.gov/
- http://www.quora.com/Where-can-I-find-large-datasets-open-to-the-public
- http://www.inside-r.org/howto/finding-data-internet
- https://www.edsurge.com/n/2014-01-21-education-datapalooza
- 1,001 Data Sets

[1] Tidy data sets are data sets that have a particular structure. Read more about tidy data in Hadley Wickham's paper, http://vita.had.co.nz/papers/tidy-data.pdf