

# Introduction

For the final project, you will create an **explanatory** data visualization from a data set that communicates a clear finding or that highlights relationships or patterns in a data set. Your work should be a reflection of the theory and practice of data visualization, and your final deliverable will be a write up along with a Tableau Public workbook.

We will provide some options of data sets to explore; however, you may choose to explore an entirely different data set. You should be aware that finding your own data set and cleaning it using Python, R, or some other language can take considerable time and effort. This can add as much as a day, a week, or even months to your project so embark on the adventure to find and clean a data set if you are truly prepared with data wrangling skills.

You have four options for this project. You should pick an option based on your prior experience with data munging and exploratory data analysis. The option you choose will not affect the evaluation of the project.

**Where to View Dataset Options** You can view the Dataset Options:

In Step One below further down this same classroom webpage currently being viewed

Download the list of Dataset Options and View as PDF: Click the "Resources" tab in the leftmost panel of your classroom, click the file name "Data\_Set\_Options\_Project\_Create\_a\_Tableau\_Story" to download a PDF with the options listed.

Download .csv files of the Datasets themselves from the download links below, in the doc, or from the Resources tab.

- **Option 1:** Select one of the beginner data sets, which already has a summary of findings. Then, create a visualization that communicates the findings.
- **Option 2:** Select one of the intermediate data sets. You will investigate the data set to share a story or message about the data and then create a suitable visualization.
- **Option 3:** Find a data set, investigate it, and share your findings in a visualization. Your final graphic should primarily be explanatory, but it may also contain exploratory components. You can find a list of recommended websites to find data sets in Step One below. You should be aware that finding your own data set, cleaning the data set, and analyzing it (using R, iPython Notebook, or another tool) can take considerable time and effort. This can lengthen the time you spend on your project by days, weeks, or even months. Choose the option only if you feel prepared for a challenge!

Now, on to the details!

## Step One - Choose a Data Set

First, you will choose a data set or find a data set to explore and visualize. You should choose a data set based on your prior experiences in programming and working with data. The data set you choose will not increase or decrease your chances of passing this project.

### Data Set Options - Project: Create a Tableau Story

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

#### Beginner - [Baseball Dataset](#)

Overview: 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. Notes: Create a visualization that shows differences among the performance of the baseball players.

#### Intermediate - [Flights Dataset](#)

Overview: 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](#).

Notes: 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 - [Prosper Loan Dataset](#)

##### [Prosper Data Dictionary](#) Explains Variables

Overview: 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 income, and many others.

Notes: Ask your own questions about this data set to find interesting trends in the data.

#### Advanced - [PISA Dataset](#) The unzipped PISA Data csv file is 2.75 GB.

##### [PISA Data Dictionary](#) Explains Variables

Overview: 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](#).

If you want to know more about the survey design, the details can be found in the technical report [here](#).

Notes: 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.

Choose Your Own Dataset - Difficulty 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<sup>1</sup> (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: [Pew Global](#) - [World Bank Data Bank](#) - [Quora Post inside-r](#) - [Data Palooza](#)

<sup>1</sup> 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>

Note: [Google Doc Link Dataset Options List](#) Alternative way to view identical list as above. Not usable on certain Networks.

## Step Two - Get Organized

Eventually you'll want to submit your project and share it. To do so, you need to create a zip folder that includes the following:

- **Write-up:** PDF or Markdown file that includes links to your Tableau Public workbooks, published online, and a write-up with four sections. See [HERE](#) if you need help publishing your Tableau Public Workbook.
  - **Summary:** in no more than 4 sentences, briefly introduce your data visualization and add any context that can help readers understand it
  - **Design:** explain any design choices you made including changes to the visualization after collecting feedback
  - **Feedback:** include all feedback you received from others on your visualization from the first sketch to the final visualization
  - **Resources:** list any sources you consulted to create your visualization

- **Data Files**
  - the final data set used to create the visualization (usually .csv, .tsv, or .json file)
  - a codebook or other files related to the data set (description, readme, license)

### **Step Three - Find a Data Story**

Explore your data set and craft a message or story around your data! Think about the overall message you want to convey and think about the comparison(s) or relationship(s) you want your readers to see. Remember that you will ultimately need to create a visualization that is explanatory, helping lead a reader to identify one or more key insights into the dataset. Feel free to use whatever visualization and data analysis tools you feel comfortable with using at this point in the process.

### **Step Four - Create Your Visualization**

First, sketch ideas for your visualization. Once you settle on a sketch, explain any design choices in that sketch, such as chart type, visual encodings, and layout, in the *Design* section of the write-up. Then, create your visualization using Tableau. The visualization must include animation, interaction, or both. See the [Project Rubric](#) for more information.

### **Step Five - Get Feedback**

Share your visualization with **at least one other person** and document their feedback. There are many ways to get feedback, and more feedback is generally better! Here are some options.

- Share your visualization with others in person and have them think aloud as they read and explore the graphic so you can document what stands out to them and how they interpret the graphic.
- Share a link to your project in your Study Group, or Slack community, depending on what version of the program you are in, and ask others to share their constructive criticisms. Be sure to offer advice to others who are seeking feedback too!

You might need to ask specific questions to prompt the reader. Here are some questions to help you. You can, of course, ask others.

- What do you notice in the visualization?
- What questions do you have about the data?
- What relationships do you notice?
- What do you think is the main takeaway from this visualization?
- Is there something you don't understand in the graphic?

### **Step Six - Document Feedback and Improve the Visualization**

For each person that gives you feedback, add the person's feedback to your write-up file in the *Feedback* section. As you improve and iterate on your visualization, update the visualization **AND** describe any changes in the *Design* section of the write-up.

You should save multiple versions of your data visualization after you make changes to it. Remember to save related files with similar numbers.

When should you save your files? You should save your files whenever you have a working version of your data visualization. If you get feedback and make changes, then create a new file and save the file as another version.

Your goal is to build evidence that you have shared your visualization, received feedback, and responded to that feedback. **You need to submit an initial version of your data visualization and the final version with the corresponding write-up.**

## **Step Seven - Review**

Use the [Project Rubric](#) to review your project. If you are happy with your submission, then you're ready to submit your project. If you see room for improvement, keep working to improve your project!