# CHAPTER 12 PROJECT

Telling a Story with Data

As you have learned throughout this chapter, data can be acquired in a variety of ways and then cleaned up, analyzed, and used to tell a story. Whether the story told about the data is true or is strongly backed by the data is something that should always be considered. While visualizations of data can help simplify the story that is being told, these visualizations can often be misleading. Consider the following data set, which shows the youth literacy rate, poverty level, access to electricity, birthrate, and mobile cellular subscriptions worldwide over a 15-year span.

Year	Youth Literacy Rate (%)	Poverty Level (%)	Access to Electricity (%)	Birthrate (per 1000 people)	Mobile Cellular Subscriptions (per 100 people)
2003	87.882	24.7	80.009	20.859	22.22
2004	88.343	22.9	80.16	20.703	27.292
2005	88.475	21	80.161	20.57	33.766
2006	88.764	20.3	81.251	20.422	41.564
2007	88.667	19.1	82.205	20.341	50.263
2008	89.451	18.4	82.284	20.232	59.375
2009	89.523	17.6	82.765	20.036	67.496
2010	89.567	16	83.3	19.809	76.162
2011	89.783	13.9	82.115	19.628	83.716
2012	90.337	12.9	84.746	19.51	87.929
2013	90.674	11.4	85.031	19.298	92.461
2014	90.993	10.7	85.553	19.21	96.053
2015	91.03	10.1	86.579	18.957	97.421
2016	91.447	9.7	87.73	18.942	100.72
2017	91.629	9.3	88.617	18.621	102.869

Source: The World Bank|Data, accessed December 18, 2021, https://data.worldbank.org/.

1. Suppose the following graph was given as part of a presentation on world youth literacy rates. What does the graph seem to imply?

Youth Literacy Rate versus Access to Electricity
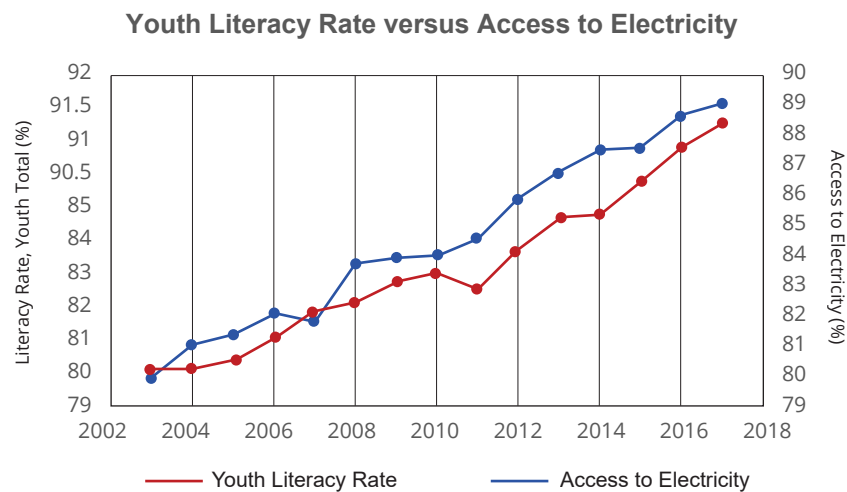
2. Explain how the graph might be misleading.

3. Create a graph that shows the data graphed using the same vertical scale. Does this graph change the story that is told by the two data sets? If so, describe how. If not, explain why.

4. Find the Pearson correlation coefficient between the youth literacy rate and access to electricity rounded to nearest ten thousandth. With a level of significance of $\alpha = 0.01$, is the relationship statistically significant?

5. Based on your findings in part 4, if you wanted to show the relationship between youth literacy rates and access to electricity, would you use the graph from part 1 or the graph from part 3 in a presentation? Which graph tells the better story about the connection? Explain your reasoning.

6. Calculate the Pearson correlation coefficient (rounded to the nearest ten thousandth) for the youth literacy rate compared to each of the remaining categories (that is, compared to poverty level, birthrate, and mobile cellular subscriptions). Determine which of these relationships are statistically significant at a level of significance of $\alpha = 0.01$.

7. Using the data provided in the table, create an infographic illustrating how youth literacy rates are related to poverty level, access to electricity, birthrate, and mobile cellular subscriptions. Your infographic should contain at least 4 different elements that clearly explain the relationships between youth literacy rates and each of other categories of the data set. (**Hint:** Search the internet for infographics to get ideas on how to structure your infographic.)

8. Is the infographic you created intentionally misleading in anyway? If so, explain how. Explain your reasoning for including misleading information.