Building Dashboards with Tableau

Data visualization appears in many forms, including line graphs, bar charts, pie charts, and scatter plots. Cady emphasized that data visualization is essential in any data science deliverable (2017). While data visualization is an ancient approach to understanding data, Baker wrote, “what's changed are the kinds and size of data that can be represented this way, and the many more sophisticated ways in which you can show it and share it” (2018).

Multiple related data visualizations can be brought together in a dashboard, further increasing the sophistication with which data can be explored. Dashboards may have fixed visual data elements that cannot be changed by the user (such as a company’s stock ticker). Dashboards also typically have interactive data visualizations that users can manipulate to explore data. For example, a dashboard could include interactive charts that provide a way for business leaders to understand sales information by drilling down to see sales performance by salesperson, geographic region, marketing campaign, product, or other variables. By using dashboards to explore data, business leaders and knowledge workers can ask and answer questions. The consequent business insights can be used to guide business decisions (such as adjusting production goals to meet a growing market demand or reacting quickly to production issues).

Developing the Dashboards

At their best, dashboards provide a streamlined and intuitive means by which business users can effectively and easily explore complex data. Cotgreave (2017) provided a few key guidelines for the development of effective dashboards. In particular, Cotgreave recommended using a grid layout, using fonts consistently to reflect hierarchy, simplifying the use of color, avoiding clutter, and drawing visual attention to key findings. In this project, Tableau Desktop

Public Edition, a popular tool for building dashboards, was used to explore data from the Robert Wood Johnson Foundation's US County Health Rankings (2014). This dataset contains rate averages for a variety of health measures (such as adult obesity, childhood poverty, and physical inactivity). Rate averages are reported geographically (at county, state, and national levels) and by year spans (in averages representing one to three years). Since the original dataset contained one row of data for every unique combination of measure, location, and year span, the first step was to convert the data to a more readable format. A "crosstab" chart was generated within Tableau, with years serving as column headings and measures and locations represented in rows. Based on the review of the crosstab chart, two interactive dashboards were conceptualized to support exploring healthcare measures. The first dashboard was sketched to show averages at state and regional levels. The second dashboard was sketched to show temporal trends for averages by state.

After the dashboards were conceptualized, sheet were configured in Tableau to draw the necessary data visualizations. The Tableau interface for creating data visualizations is intuitive. It includes a wide range of charts and graphs and it provides dynamic guidance throughout the design process. Tableau provides numerous customizations for each of its charts and graphs. For example, trend lines can be added to graphs to show median or mean values. The lines can be drawn with or without confidence bands. Symbols on charts can represent data values using size, shape, or color. Tableau provides default features (such as "tooltip" text) to simplify the appearance of charts and graphs while also providing a way for users to easily access additional information about specific data points. While tooltips and labels have default layouts, developers are able to customize their content based on business needs.

For the current project, some data visualizations required the aggregation of states into regions. To support this aggregation, a calculated field (see Appendix A for code) was created to assign regions based on state using U.S. Census regions (U.S. Census Bureau's 2017 trends data documentation, 2017). In the original Robert Wood Johnson Foundation US County Health Rankings dataset (2014), some measures reported averages for single years, but other measures reported rolling averages over two or three years. As well, the measures that were reported each year varied. For example, only Premature Death rates were reported before 2002. In 2002, Children in Poverty and Sexually Transmitted Infections rates were added. Gradually, additional measures were added (see Figure 1).

	Year span								
Measure name	1999-2001	2000-2002	2001-2003	2002	2002-2004	2003	2003-2005	2004	2004-
Adult obesity								0.253	
Children in poverty				0.180		0.190		0.190	
Daily fine particulate matter						12.192		11.956	
Diabetic screening							0.801		
Mammography screening							0.640		
Physical inactivity								0.259	
Premature Death	8,338.046	8,391.117	8,445.645		8,471.900		8,494.508		8,2
Preventable hospital stays							91.389		
Sexually transmitted infections									
Unemployment				0.057		0.060		0.057	
Uninsured									
Violent crime rate							281.104		2

Figure 1. Crosstab chart showing rates by year span for a variety of health measures, averaged over geographic areas. Because of the differences in how years are represented, and because not all measures were reported for each year, there are numerous missing values in the crosstab chart. Data from Robert Wood Johnson Foundation US County Health Rankings dataset (2014).

One data visualization required for the dashboard showed a map of the United States drawn using color intensity to represent a measure's rate in each state. Since the map would be managed by a page to show trend data dynamically, it would only be used with the healthcare

measures that were reported annually (rather than those that were reported in two- or three-year spans). Filters were added to the map to select only single years and to select only state-level data. Puerto Rico, Alaska, and Hawaii were excluded so that the map would represent only the continental United States. On the dashboard, the filter selection was changed from "all values in the database" to "only relevant values" so that only relevant measures appeared as choices. The page feature was simple to use and provided an intuitive way to dynamically view changes in data over time.

It was relatively straightforward to create the Tableau dashboard using drag-and-drop and clicking functionalities. Horizontal and vertical objects were used to position data elements in a grid, as recommended by Cotgreave (2017). Tableau's automatic integration of data elements simplified design (for example, when a healthcare measure was selected from the dashboard filter, both related charts were automatically updated). Once visualizations appeared in dashboards, they were adjusted several times in an effort to meet Cotgreave's key guidelines for the development of dashboards (2017). After the two dashboards were finalized, they were added to a storyboard.

Using the Dashboards for Analytics

The first dashboard, "Measures by Region," shows state rate information about a single selected measure (see Figure 2). At the top of the dashboard, a selection tool allows the user to select the healthcare measure. Following Cotgreave's recommendation for drawing attention to key data points (2017), the highest and lowest rates for the selected measure are displayed in the upper left corner with state abbreviation and rate in text. The top and bottom rates are also represented visually using a packed bubble chart. Below, a scatter plot shows rate comparisons

across regions. The scatter plot is overlaid with box plots to support visual analysis regarding difference between regions. In the right-hand column, a bar chart shows the rates for all states, sorted highest to lowest. Cotgreave (2017) recommended simplifying color. For this dashboard, the color for all visual data shares the same color legend based on region. Initially, color choice for regions was automatically selected. However, since the packed bubble chart had only two data points, it used a different color key compared with the scatter plot and bar chart. Fortunately, the problem was easily corrected by assigning colors for each region. Colors were selected from the built-in palette designed to accommodate people with color blindness.

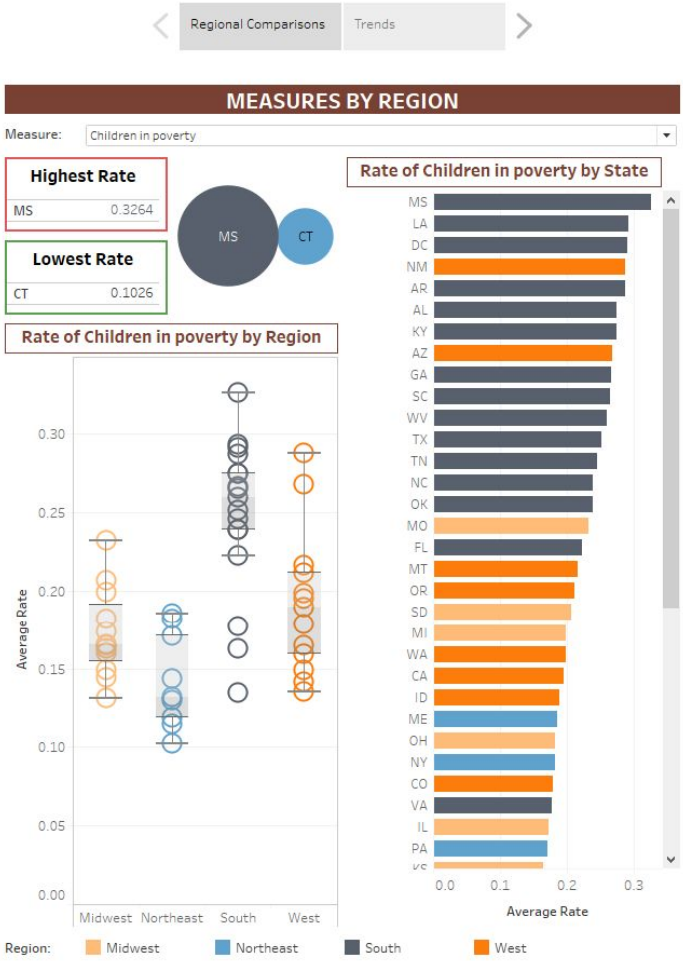

Figure 2. Dashboard showing regional comparisons.

Exploration of the Regional Comparisons dashboard showed that there are significant differences between regions for several measures. For example, boxplots representing regional data points for children in poverty are far apart, suggesting real differences in childhood poverty rates among U.S. regions (see Figure 2). Data visualization also revealed other patterns. For example, rates for diabetic screening were more similar across states and regions compared with other measures (compare Figure 2 with Figure 3).

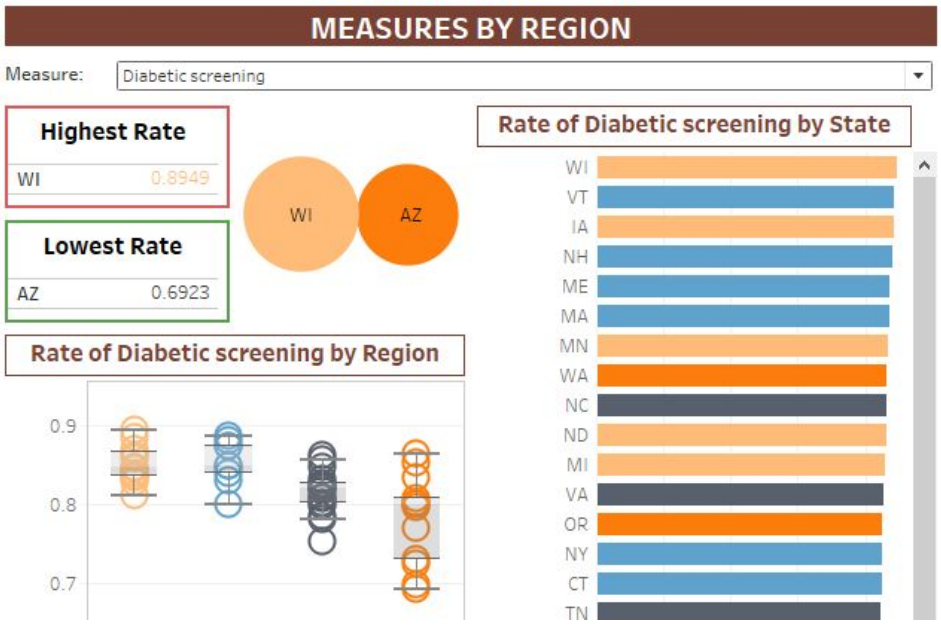

Figure 3. Diabetic screening data had less variation across states and regions compared with other measures.

The second dashboard, "State Trends for Measures," shows state rate information for a single selected measure plotted with time (see Figure 4). On this dashboard, a single-color gradient palette was selected to highlight differences in relative rates. Consistent with the first dashboard, the "State Trends for Measures" dashboard has a filter to select measure. Below the filter selection tool, the dashboard has two adjacent data visualizations. On the left, a map

represents relative rates between states for a given year. A page allows the user to change the year that is represented on the map. On the right, a scatter plot shows all annual rates for the measure using a dot for each state. Dots are colored using the same gradient to represent relative rate values. On the scatter plot, trend lines with 95% confidence intervals are displayed for each year. These trend lines help the user to see broader national trends in the data.

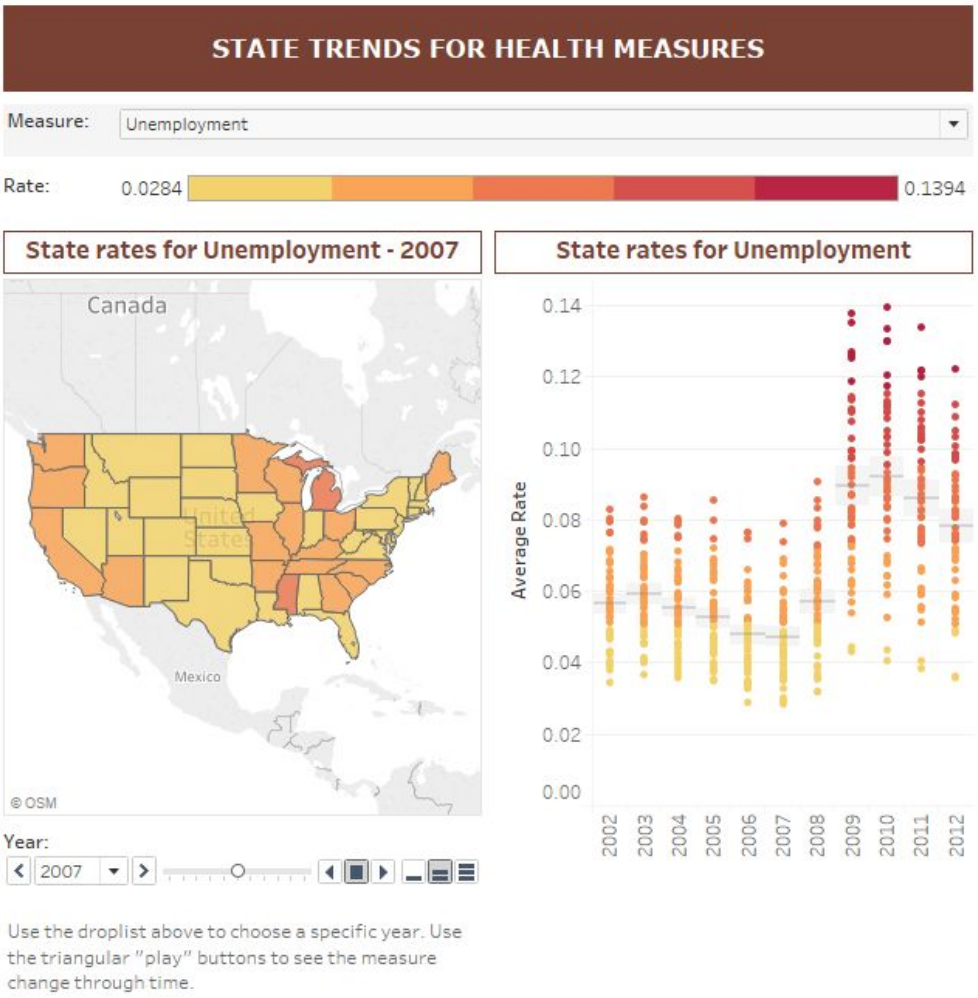

Figure 4. State trends for health measures.

The State Trends for Health Measures dashboard was useful in revealing several trends over time. For example, an economic downturn in 2008 is strongly reflected in the data

visualization for unemployment (see Figure 4). Not surprisingly, rates for children in poverty also reflect the economic downturn. As well, while there is variation in adult obesity data, the trend lines show that the nationally mean for adult obesity rate increased each year (2004-2010).

Conclusion

Tableau is a powerful tool for understanding complex data. Tableau contains numerous features that support developers in developing compelling data visualizations, thereby helping users to discover data trends and patterns. As well, Tableau features, including tooltips and filters, allow the user to explore data on a finer scale. While Tableau is intuitive to use, and it creates impressive graphics, it is also so feature rich that that it is important to use training resources to have a better understanding of all the features the tool has to offer.

While Tableau's calculated fields are useful, there are circumstances where it would likely be simpler to adjust the data prior to bringing it in as a data source in Tableau. For example, in the current application, the dataset raw values represented rates set to different scales depending on measures. While most healthcare measures reported percentage rates (from 0 to 1), others reported whole numbers that represented the number of cases per 1,000 or 100,000 in the population. For this project, rates were used as reported. In a production environment, the dashboards would have been improved with a dynamic footnote explaining the data for the selected measure. If business intelligence usage required the comparison of data across measures, creating scaled rates could be more performed outside Tableau. For example, the R Foundation's R Version 3.5.1 software could be programmed to read the data file, add a field representing scaled rates, and save the updated dataset as a comma separated values (CSV) file. The updated data file could then be used as the data source in Tableau to generate data visualizations.

Similarly, a program such as R could roll data reported in single year spans into two- and three-year span averages in order to facilitate temporal comparisons between measures reported with different year spans.

References

2017 trends data documentation. (2017). County Health Rankings [PDF]. Retrieved from http://www.countyhealthrankings.org/sites/default/files/2017TrendsDocumentation.pdf

Baker, P. (2018, December 14). The Best data visualization tools for 2019. Retrieved from https://www.pcmag.com/roundup/346417/the-best-data-visualization-tools

Cady, F. (2017). *The Data science handbook*. Hoboken, NJ: Wiley.

Census regions and divisions of the United States. Retrieved from https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf

Cotgreave, A. (2017, October 19). 7 tips and tricks from the dashboard experts. Retrieved from https://www.tableau.com/about/blog/2017/10/7-tips-and-tricks-dashboard-experts-76821

Frye, C. (2017, May 10). Tableau 10 essential training [Lynda.com online course]. Retrieved from https://www.lynda.com/Tableau-tutorials/Tableau-10-Essential-Training/5005402.html?srchtrk=index%3a14%0alinktypeid%3a2%0aq%3avisualization+tools%0apage%3a2%0as %3arelevance%0asa%3atrue%0aproducttypeid%3a2

Porter, J. Principles of user interface design. Retrieved from http://bokardo.com/principles-of-user-interface-design/

Robert Wood Johnson Foundation. 2014. US County Health Rankings [data file]. Retrieved from https://public.tableau.com/s/sites/default/files/media/County_Health_Rankings.csv

Tableau Desktop Public Edition. (2018). Public [Software]. Available from https://public.tableau.com/en-us/s/

The R Foundation. (2018). R (Version 3.5.1) [Software]. Available from https://www.r-project.org/

Appendix A

Code for Region Calculation

The region was computed with a calculated field using the code below.

```
IF ((ATTR([State])="CT") OR

(ATTR([State])="ME") OR

(ATTR([State])="MA") OR

(ATTR([State])="NH") OR

(ATTR([State])="RI") OR

(ATTR([State])="VT") OR

(ATTR([State])="NJ") OR

(ATTR([State])="NY") OR

(ATTR([State])="PA"))

THEN "Northeast"


ELSEIF ((ATTR([State])="IN") OR

(ATTR([State])="IL") OR

(ATTR([State])="MI") OR

(ATTR([State])="OH") OR

(ATTR([State])="WI") OR

(ATTR([State])="IA") OR

(ATTR([State])="NE") OR

(ATTR([State])="KS") OR
```

```
(ATTR([State])="ND") OR
(ATTR([State])="MN") OR
(ATTR([State])="SD") OR
(ATTR([State])="MO"))
THEN "Midwest"

ELSEIF ((ATTR([State])="DE") OR
(ATTR([State])="DC") OR
(ATTR([State])="FL") OR
(ATTR([State])="GA") OR
(ATTR([State])="MD") OR
(ATTR([State])="NC") OR
(ATTR([State])="SC") OR
(ATTR([State])="VA") OR
(ATTR([State])="WV") OR
(ATTR([State])="AL") OR
(ATTR([State])="KY") OR
(ATTR([State])="MS") OR
(ATTR([State])="TN") OR
(ATTR([State])="AR") OR
(ATTR([State])="LA") OR
(ATTR([State])="OK") OR
```

```
(ATTR([State])="TX"))

THEN "South"


ELSEIF ((ATTR([State])="AZ") OR

(ATTR([State])="CO") OR

(ATTR([State])="ID") OR

(ATTR([State])="NM") OR

(ATTR([State])="MT") OR

(ATTR([State])="UT") OR

(ATTR([State])="NV") OR

(ATTR([State])="WY") OR

(ATTR([State])="AK") OR

(ATTR([State])="CA") OR

(ATTR([State])="HI") OR

(ATTR([State])="OR") OR

(ATTR([State])="WA"))

THEN "West"

END
```