VISUALIZING DATA WITH TABLEAU

A DATA DOZEN WORKSHOP

BUILT FOR THE STRATEGIC DATA PROJECT
2020 WINTER WORKSHOP

DATA DOZEN
DATA VISUALIZATION EDUCATION

LET'S GET STARTED

This engaging, two hour workshop will demystify Tableau Software, introduce key data visualization fundamentals, and jumpstart your ability to create interactive dashboards with your data. With an emphasis on Tableau's key functionalities and practical applications, we'll progress from basic charts to dynamic dashboards that incorporate interactive filters and design theory. We'll use a data set about the admission process at American universities and Tableau Public (which is completely free!) to explore the very best of what this data visualization tool has to offer.

WELCOME

- Personal introductions: what's your name, where do you work, and what are you hoping to learn?
- Laptop Check: Be sure you've downloaded Tableau Public and download the data set.
- Our agenda has three major sections:
    - Data + Tableau
    - Warm Ups
    - Dashboard 1
- We have three primary goals today:
    - Play, alot. This is like art class. Your mug will not come out of the kiln looking identical to your neighbors'.
    - Set you up to keep learning. You won't be a jedi master at the end of the class, but you'll be on your way.
    - Breakdown the qualitative v. quantitative division. Lean into the direction that's new to you.
- Why Tableau
    - Doesn't require coding
    - Exploratory and explanatory
    - Dynamic Output
    - Free Version

HELPFUL LINKS

- Companion Website
  We move quickly during the workshop. There's a lot to cover in just two hours. If you miss a step or link, don't worry. Everything we cover today is available in the workshop's companion website.
- Tableau Workbook
  A complete version of the sample workbook you're seeing today is available here. You can download your own copy and deconstruct each dashboard and sheet.

DATA + TABLEAU

WHAT MAKES A VISUALIZATION

When we distill visualizations down to their basic form, we get two things: some type of mark and some way to tell multiple marks apart.

- There are three basic types of marks we use to represent our data:
  - Points: scatter plots, box plots, dot maps
  - Lines: bar chart, histograms, line chart, gantt chart
  - Area: treemap, area chart, filled maps,
- Then we tell these marks apart by leveraging:
  - Position: up, down, left, right
  - Size: length, width, angle
  - Color: categorial, saturation
  - Shape: icons, polygons

CONNECT TO DATA

Before we can build any visualizations, we need to connect our data to Tableau. (Well actually before that, we need to gather and structure our data, which is easily 95% of the work, but we'll set that aside for today and enjoy the last 5%.) We will open a new Tableau Public workbook, connect it to the data set we've downloaded, and then go to our first Tableau worksheet.

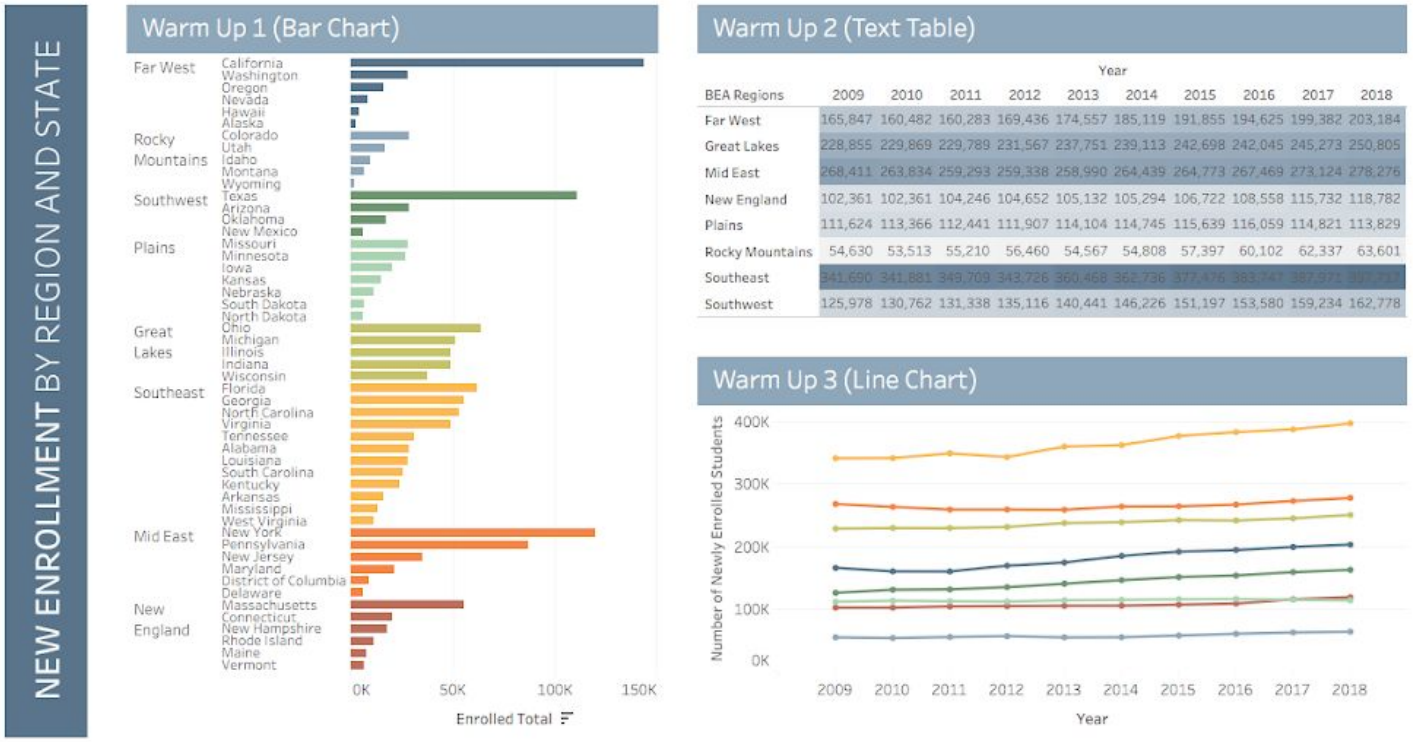
- Connect to Data
  - Launch Tableau Public.
  - Under Connect and To a File, click Microsoft Excel.
  - Find the "Tableau Data - Harvard SDP Winter Workshop 2020" data set you've downloaded and click Open.
  - Drag the Table 1 - Clusters sheet onto the orange Drag Sheets Here text.
  - The data will load in the preview area.
- Go to Worksheet
  - Click the orange tab labeled Sheet 1 to go to Tableau's building environment.

TABLEAU WORKSHEET ENVIRONMENT

Welcome to Tableau's worksheet environment! This is where we build all our visualizations. On the left side of the screen, you'll see all of your variables are now data pills sitting in the data pane. There are three things to note about these data

pills and three places we'll talk about dragging and dropping them in order to create visualizations. (If the Show Me menu is showing on the right hand side, just click the words to collapse the menu. That will make everything easier to see.)

- Data Pane
  - Dimensions vs. Measures: Tableau has split up your variables into dimensions, which are descriptive pieces of data, and measures, which are the numbers that you would do math on.
  - Discrete vs. Continuous: You'll notice some of the data pills are blue, indicating they're discrete pieces of data, think categorical boxes, as opposed to the green pills, which represent continuous data, or data that exists on some sort of pre-existing spectrum, like numbers do.
  - Data Types: Tableau has made it's best guess about each variable's correct data type: numer (decimal), number (whole), date and time, date, string, or geographic data.
- Where to Drag-and-Drop
  - Columns and Rows: Placing data pills here function as your typical x and y axes.
  - Filters Card: Placing data pills here allows you to filter your visualization by different variables.
  - Marks Card: This is Tableau's secret sauce. By adding variables to color, size, text, detail, and tooltips, Tableau enables you to layer additional information on top of your standard x and y axes.
- Data Pane Clean Up
  - Data Type
    - Select the Telephone variable that has been classified as a measure and drag it up to be classified as dimensions. (We'll never do math on variables like that, just like we wouldn't do math on a student ID.)
  - Group By
    - Click the dropdown menu to the right of Dimensions and change the Group By function from Data Source Order to Group By folder.
    - Select all the measures related to the ACT, right click on the highlighted measures, select Folders, and then Create new folder. Name it ACT, and then collapse all of those variables. You can use this method to clean up the rest of your data pane however you'd like. You could create folders for the SAT variables, the admissions variables, the location variables in the dimensions section, etc.
  - Number of Records check
    - Drag Number of Records to text on the marks card.
    - You should have 16,580 records (1,658 universities with 10 rows each for admissions data from 2009-2018.)



WARM UP 1: BAR CHART

Let's start building! We'll get to know Tableau's environment by warming up with a bar chart. Notice how we use each key piece of the Tableau environment: the rows and columns cards, the filters card, and the marks card.

- Rows and Columns
  - Drag Region to Rows.
  - Drag Enrolled Total to Columns.
  - Hover over the X axis and click the sort icon once to arrange the bars from biggest to smallest.
- Filters Card
  - Look at the bottom left hand corner of the screen. The sum for this visualization is nearly 15 million newly enrolled students. What's wrong with this? (Hint: the total students enrolled in US institutions is roughly 20 million.)
    - Drag Year to Detail on the marks card. All of the years are added together.
  - Drag Year to Filters, click Year, and select 2018.
    - Click the dropdown menu on the Year data pill on the Filters Card and select Show Filter.
    - Click the dropdown arrow on the filter that just appeared and change the filter type to Single Value Slider.
    - Click the dropdown menu again, select Customize, and deselect Show "All" Value.
- Hierarchy

    - Right click on BEA Regions in the data pane, click Hierarchy, then Create hierarchy…
    - Name the hierarchy "Region, State". A little hierarchy symbol will appear in the data pane.
    - Drag State into the hierarchy by placing it just beneath Regions. A black line will appear. Once it does, release.
    - Now look back at your bar chart. A small plus sign has appeared on the BEA Regions on the rows column. Click on the plus sign to expand it. Click on the minus sign to collapse it.
- Marks Card
    - Color
        - With the states showing, drag BEA Regions on to Color.
        - Click on Color, then Edit Colors… to change the colors. In the example, I used the Miller Stone color palette.
        - Click Assign Palette, or click on each region in the list to assign the colors one at a time.
    - Tooltips
        - Click on Tooltip on the marks card to edit it. Instead of a list, we'll make a sentence.
        - In <YEAR(Year)>, <State> had <SUM(Enrolled Total)> newly enrolled students in its universities..
        - We'll make the font 14 and black, and only bold the important variables that change (State and Enrolled Total.)
- Name Your Sheet
    - Double click on the bottom tab where you see Sheet 1 and name your sheet to something easy to recognize, like Warm Up 1 - Bar Chart.

WARM UP 2: TEXT TABLE

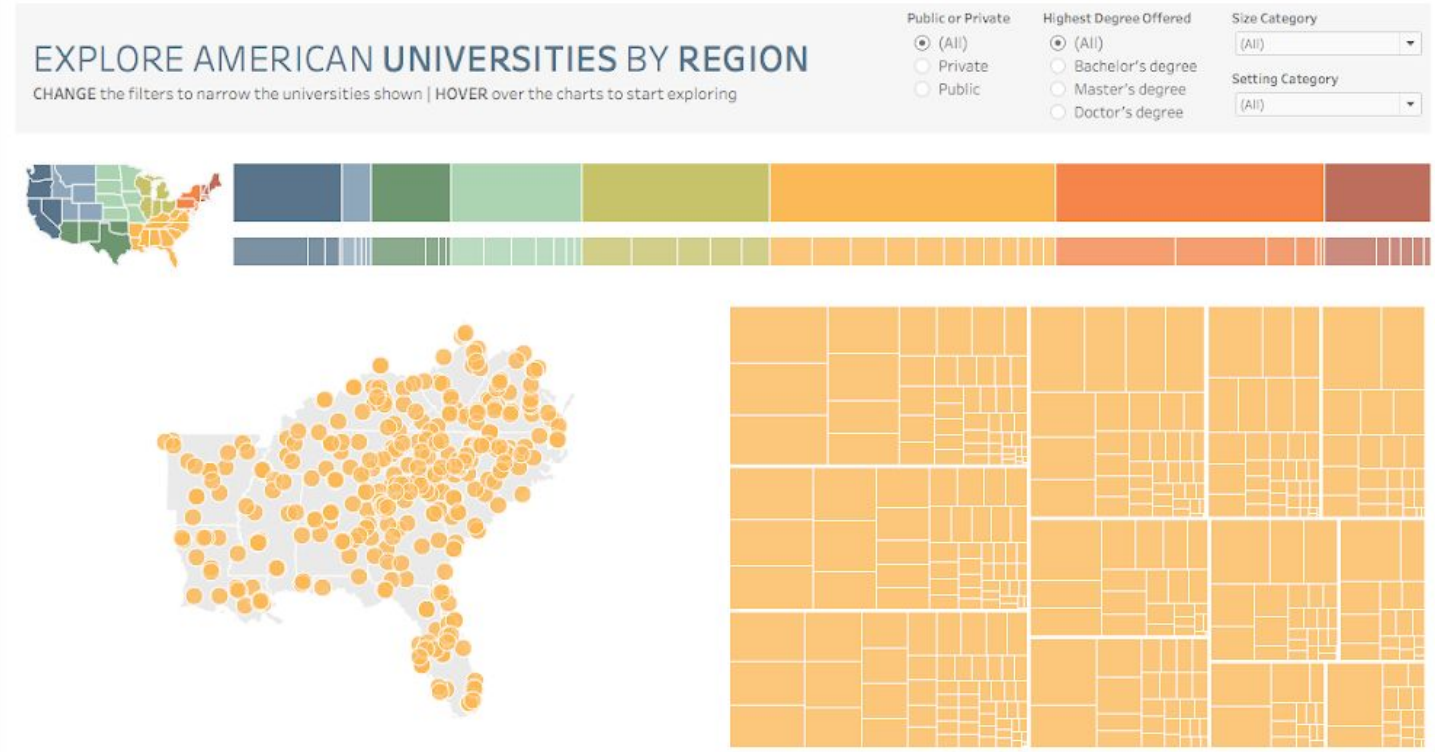
Often we still need to show our data as a text table. Let's build one here using the same data pills we used in our bar chart.

- Text Table
    - Create a new worksheet
    - Year to columns
    - BEA Region to rows
    - Enrolled Total to Text on the marks cards
    - Change the view dropdown menu from Standard to Fit Width
    - Note that your hierarchy on Region is still available for you.
- Highlight Table
    - Enrolled Total to Color
    - Change the marks dropdown menu from Automatic text to square
- Change Continuous Color
    - Click on Color and select Edit Colors…
    - You'll notice that now that we have a continuous measure on color instead of a discrete color, we now have continuous color palettes to choose from.
    - You can select a new palette from the dropdown menu, or click the dark blue square and select your own baseline color from which to build a continuous palette.
    - Name your sheet

WARM UP 3: LINE CHART

While the text table might be helpful in some cases, it's often more useful for the human eye to see a chart alongside it. Let's take these same pills and create a line chart with them. You'll see how much more immediate information you have by glancing at this line chart compared to glancing at the previous text table.

- Line Chart
  - Create a new worksheet
  - Year to Columns
  - Enrolled Total to Rows
  - Region to Color
  - Click on Color and select the Markers option with dots
- Hierarchy
  - Notice that the BEA Regions pill on the marks card still has that hierarchy plus sign available to you. Click on it to expand the level of detail from Region to State.
  - While the majority of the state's lines become illegible, four states pop up to the top immediately: California, New York, Texas, and Pennsylvania. While the first three states are known to have the largest populations in the country, Pennsylvania is a surprisingly heavy hitter given its smaller population.
  - Name your sheet



DASHBOARD 1

This dashboard enables our audience to easily explore universities in the contiguous United States by region. The two 100% filled bar charts show us what percentage of universities are located in which regions along the top, and that same breakdown by state just below it. The map legend of the US allows us to see which states are included in which region. The individual institutions are displayed on the region map, and each institution's 2018 entering class is displayed on the treemap to its right. We're quickly able to see which institutions have the largest entering class in the state, and which have the smallest. Finally our dashboard actions create a very interactive experience for the end user. They can filter the schools they see by clicking on a state or a region, and can adjust the type of schools they want displayed using the filters above.

MAP COLOR LEGEND

We'll start by building our map color legend. The rest of the color palette for the entire dashboards will hinge on this, so we'll take the time to get our colors just right.

- Filled Map
  - Create a new sheet
  - Double click Latitude and Longitude
    - Ask yourself, what does that mark represent?
  - State on detail on the marks card

- Change the mark type from the automatic circle to the Map option
- From the menu bar, choose Map > Map Layers and pull the washout slider completely to the right.
- From the menu bar, choose Map > Map Options and uncheck all options. This will keep the map from shifting accidentally when you're end user is working with it.

- Set
  - We want to filter our map down to just the contiguous United States. Instead of clicking all of the areas we want to exclude on each map, we'll make a set that just keeps the 48 states plus D.C. we want to display here.
  - Click the dropdown on the State data pill in the data pane and select Create and Set.
  - Rename your set from Set 1 to "Contiguous United States"
  - Click All and note that you have 54 values in the set.
  - Uncheck Alaska, Guam, Hawaii, Puerto Rico, Virgin Islands. We're down to the values we want to keep.
  - Click OK and see a new section appear in your data pane with your set.
  - Drag your new set onto the filters card.
- Colors
  - Drag BEA Region to color on the marks card
  - Change the order of the regions to match the geographical order from east to west by clicking on the dropdown menu on the Region data pill in the data pane, and choose Default Properties > Sort > Sort by > Manual, and then rearrange the regions.
  - Click on Color on the marks card, and then edit colors. Change the palette to the Miller Stone palette, and assign the palette. Swap any individual colors as you see fit.
  - Click on Color again and add a white border around each mark.
  - Change the opacity from 80% to 100%.
- Tooltips
  - Drag Unit ID to tooltips and you'll see it shows up as an asterisk. That signals that there's more than one Unit ID in that mark.
  - Change the dropdown menu on the Unit ID pill on the marks card to Measure > Count Distinct
  - Click on tooltips on the marks card and edit your tooltip (Larger font, all black, only bold what is important and changes)

100% BAR CHART BY REGION

Now let's visualize the same data we've just added in our tooltip, letting folks easily see how many institutions are in each region.

- Bar Chart
  - Create a new sheet
  - Drag Unit ID to Columns and change the aggregation to Measure > Count Distinct (You'll get 1,658).
  - Drag your contiguous US set onto the filter. (Now you'll have 1,612 institutions.)
  - Drag BEA Regions to Color and add a white border around each mark.
- Table Calculation: Percent of Total

- Click the dropdown menu on the Count Distinct Unit ID pill on columns and select > Quick table calculation > Percent of Total.
- Right click on your x-axis and fix the axis from 0 to 1.
- Edit up your tooltips and name your sheet

100% BAR CHART BY STATE

Now we'll build on our work with the Regions bar chart, and create a nearly identical version at the State level.

- Duplicate
  - Right click on your regions bar chart sheet and select duplicate.
  - Rename the sheet State bar chart.
- Detail
  - Expand the hierarchy plus sign on the Regions pill on the marks card. The State pill will appear on detail.
  - Sort the State pill by the field count distinct Unit ID in ascending order.

TREEMAP

Now it's time to get down to the institution level. We'd like to see all the schools for a given region by their enrollment class. When we build this out as a bar chart, it's just too difficult to see every institution in any sort of helpful way. When we run into legibility issues like this, it's usually time for a new chart. Here a treemap will be the perfect structure.

- Treemap
  - Unit ID to Detail
  - Total Enrolled to Size
  - Filter by Contiguous US set and to just 2018 year
  - Institution Name to Tooltip (Texas A&M had the largest entering class in 2018)
  - Drag Region to Color and expand to the State level of detail using your hierarchy
  - Show the Region filter and select just one region.
- Table Calculation: Rank
  - Drag another Total Enrolled data pill to tooltips.
  - Click the dropdown menu on the pill, select > Quick table calculation > Rank. (All ranks will show up as 1.)
  - Click the dropdown menu again and select > Compute Using > Unit ID. Now we can see the largest entering classes to smallest with their rank for each state.
- Website to Detail
  - Finally, drag the website data pill to detail. We'll use that in our dashboard later.

DUAL AXIS MAP

Last but not least, we'll build our map with the individual institutions locations on it as well.

- Dual Axis Combination Chart

- Duplicate the map you made for the key and rename it Dual Axis Map
- Move the BEA Regions pill that on Color on the marks card and move it to Filters. Select just one region.
- Now change the color of the marks to a light grey.
- Drag a second Longitude pill to the columns shelf. This will duplicate the map and three marks cards will appear: one that controls both charts, one that controls the left chart, and one that controls the right chart.
- Click on the last marks card to edit the right map. Change the mark type from Map to Circle.
- Add BEA Region to color so that your dots match the regions colors.
- Replace State on detail on the marks card with Unit ID on detail instead.
- Add Website and the Institution Name to the tooltips.
- Click on the dropdown menu on the second Longitude pill on the Columns shelf, and click dual axis.

- Filters
  - Public or Private (alias)
  - Size Category (default sort)
  - Highest Degree Offered (groups)
    - Click on Highest Degree Offered in the data pane and select > Create > Group
    - Select all four doctoral degrees, right click and group them into a single category.
    - Show the filter for the Highest Degree Offered
  - Setting (groups)
    - Create a new sheet and drag Size and Setting to Rows.
    - Select all of the "Highly Residential", right click, and create a group. Rinse and repeat for the remaining settings. Then you can go back to your scatter plot and add that Setting Group to your filters.

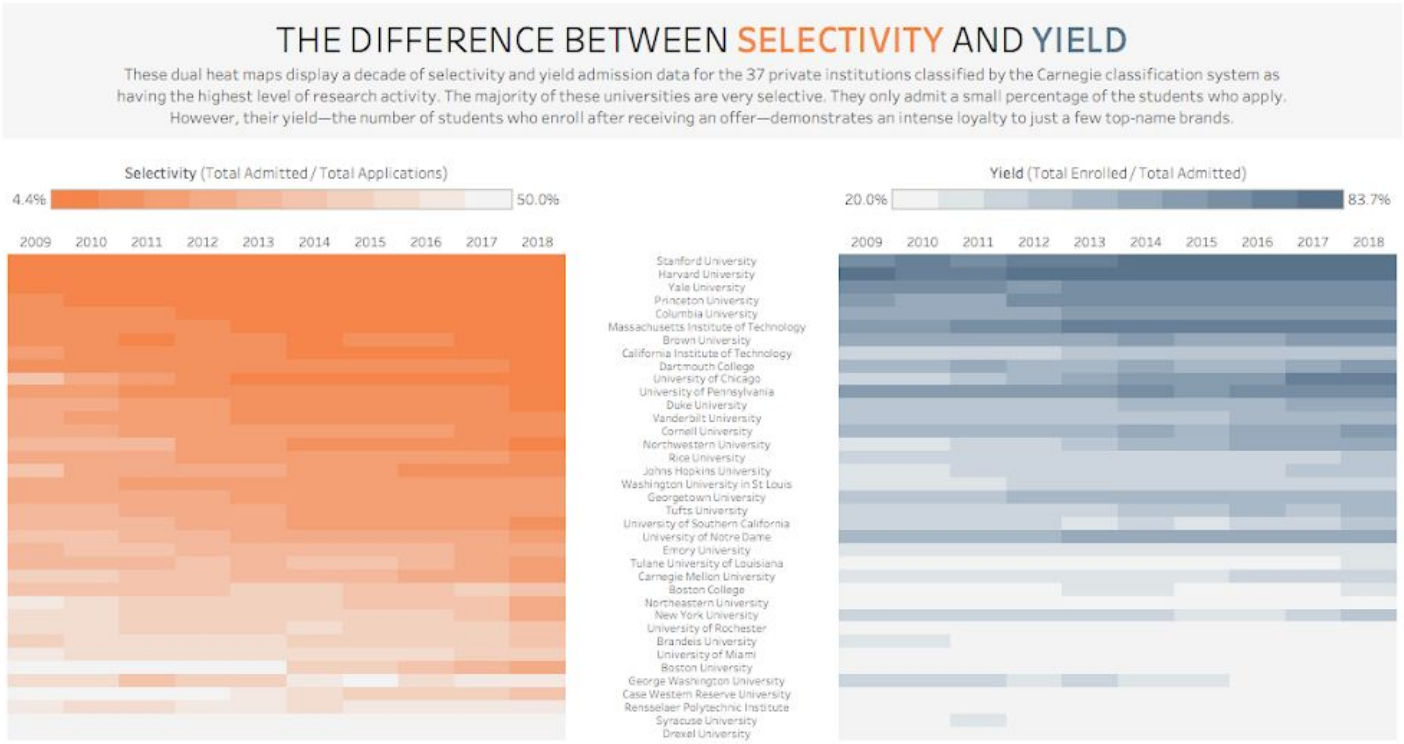
DASHBOARD 1

Now it's time to stitch our five visualizations together into a single dashboard. First we'll add all the elements onto a single dashboard, and then we'll connect all of the filters and dashboard actions to work on the

- Create a Dashboard
  - Instead of clicking on the new worksheet button, click on the new dashboard button to the right of your existing sheets. It looks like a little window pain.
  - Change the Size from a min and max format to Automatic.
- Add Charts
  - Drag all five charts onto the dashboard, placing the bar charts on top of each to the right of the map key.
  - Give your
- Eliminate Anything Unnecessary
  - Right click on any titles and hide them.
  - Remove the color legend and any other items that were brought in automatically that aren't needed for the final dashboard.
- Maximize the Real Estate

- Tweak the size of your items until you feel that you've created a balanced worksheet, providing the most space to the most important items.
- Make sure each chart has the view set from Standard to Entire View.

- Help Your Audience Explore
  - Apply each filter to all sheets on dashboard
  - Highlight Action
    - Highlight School + Treemap (on Unit ID)
    - Highlight Region + State + Map (all fields)
  - Filter Action
    - Click on Region + State + Map bar chart (leave the filter)
  - URL Action
    - Click on Treemap + Dual Map, taken to website



DASHBOARD 2

Our second dashboard offers our audience a glance into how intensely competitive applying for college can be. We'll use a pair of calculated fields and heatmaps to show the selectivity and yield percentages over the last decade for the private institutions with the highest research activity. Compared with more exploratory dashboards that encourage our audience to dive into the data themselves by adjusting a set of filters, this dashboard is an explanatory dashboard. We're capturing an interesting slice of the story we've already discovered in the data, and explaining our findings to our audience.

SELECTIVITY HEATMAP

We'll start by creating our selectivity heat map. How selective a university is comes down to how many offers of admission they make out of all the applications they received. Or more simply put, the number of offers divided by the number of applications.

- Table
  - Year to Columns
  - Institution Name to Rows
  - Carnegie Classification to Filter > Doctoral Universities: Very High Research Activity
    - This will leave a 130 universities, which you can see in the bottom left corner's summary
  - Public or Private to Filter > Private not-for-profit
    - This will leave just 37 universities

    - Change view from Standard to Entire View
- Calculated Field
    - Click the Analysis dropdown menu at the top, and select Create Calculated Field…
        - You can see all of Tableau's functions by clicking on the small arrow on the right side of the popup box.
    - Give your calculated field a name, like CALC: Selectivity
    - In the text box, type your calculation: sum([Admissions Total])/sum([Applicants Total])
    - Drag your new calculated field on to Color on the marks card.
- Sort
    - Now let's sort the table by selectivity.
    - Click the dropdown menu on the Institution Name data pill sitting on Rows, select Sort…, and then set the following in the pop up box:
        - Sort By: Field
        - Sort Order: Ascending
        - Filed Name: CALC: Selectivity
        - Click the X in the top left corner to close the box
- Color
    - The color intensity is actually flipped from what we visually want to emphasize. Click on Color, Edit Colors and make the following adjustments:
        - Change the blue to an orange color.
        - Check Stepped Color, 10 Steps
        - Check Reversed
        - Advanced>> Check End and enter .5 instead of 0.815
- Tooltip
    - First, let's make sure that our calculation is accurate.
        - Drag Applications Total and Admissions Total on to tooltip.
        - Pick a mark, grab your calculator, and divide the Admissions Total by the Applications Total and make sure that you get the same selectivity in both places.
    - Then, let's explain our logic to our audience.
        - First, go to your CALC: Selectivity in your data pane, click the dropdown menu, Default Properties, Number Format, and select a percentage format with one decimal.
        - Drag Institution Name to Tooltip. Now click on Tooltip to turn it into a sentence.
        - <ATTR(Institution Name)> received <SUM(Applicants Total)> applications in <YEAR(Year)> and offered admission to <SUM(Admissions Total)>, or <AGG(CALC: Selectivity)> of those who applied.
        - Adjust fonts and sizing as needed.
- Name Your Sheet
    - Double click the sheet's name and change it to Selectivity Heat Map.

YIELD HEATMAP

Next we'll make our yield heat map. Structurally, this chart is nearly identical. However, this time the color will be based on the university's annual yield. That is the number of students who actually enrolled out of all the folks who received an offer. Or more simply, the total enrolled divided by the total admitted.

- Duplicate
  - This time, instead of creating a new sheet, right click on your Selectivity Heat Map, and select Duplicate.
  - Change this new sheet's name to Yield Heat Map.
- Calculated Field
  - Create a new calculated field called CALC: Yield.
  - sum([Enrolled Total])/sum([Admissions Total])
  - Change the new field's default number format to a percentage with one decimal.
- Color
  - Replace CALC:Selectivity with your new CALC: Yield on color.
  - Click on Color, Edit Colors and make the following adjustments:
    - Make the color palette something blue.
    - Check Stepped Color, 10 Steps
    - Advanced>> Check Start and change .0811 to .2.
- Tooltips
  - Drag Enrolled Total onto tooltip.
  - Adjust your tooltip to read logically. For example:
  - Of the <SUM(Admissions Total)> individuals who received an offer in <YEAR(Year)>, <SUM(Enrolled Total)> students enrolled in <ATTR(Institution Name)>, for a yield of <AGG(CALC: Yield)>.

DASHBOARD 2

With our two heatmaps built as individual worksheets, it's time to combine them into a single dashboard.

- Create a Dashboard
  - Instead of clicking on the new worksheet button, click on the new dashboard button to the right of your existing sheets. It looks like a little window pain.
  - Change the Size from a min and max format to Automatic.
- Add Charts
  - Drag the Selectivity Heatmap onto the dashboard.
  - Next drag the Yield Heatmap onto the dashboard, releasing it to the right of the Selectivity heatmap.
- Eliminate Anything Unnecessary
  - Right click on the Selectivity's Institution Name list and uncheck Show Header.
  - Right click on Year on both charts and Hide Field Labels for Columns.
  - Right click on each chart's title and hide it.
- Maximize the Real Estate
  - Move the color legends above their respective charts.

    - Right click on the color legends title and use this space to title the charts:
        - Selectivity (Total Admitted / Total Applications)
        - Yield (Total Enrolled / Total Admitted)
- Focus on Your Audience
    - Add a highlight action so that when someone clicks on a school's name, it highlights their selectivity row as well as the yield.
        - Actions > Add Action > Highlight > Run on Select > Target Highlighting, Selected Field, Institution Name
    - Add a title by dragging a text box above all the heat maps. Add an explanation along with the title.
        - THE DIFFERENCE BETWEEN SELECTIVITY AND YIELD These dual heat maps display a decade of selectivity and yield admission data for the 37 private institutions classified by the Carnegie classification system as having the highest level of research activity. The majority of these universities are very selective. They only admit a small percentage of the students who apply. However, their yield—the number of students who enroll after receiving an offer—demonstrates an intense loyalty to just a few top-name brands.

SAVE YOUR WORKBOOK

The final step is to save your workbook! If you haven't already created a Tableau Public account, return to the Companion Guide website for instructions to make one. Once you've done that, here's how you save your workbook.

- Go to File, Save to Tableau Public As, and then enter your Tableau Public account credentials.
- Name your workbook.
- Edit the workbook details on your Tableau Public page. You can control whether the workbook appears on your profile page and whether someone else can download the workbook and data.
- Note that you should NEVER save anything private on Tableau Public. This is a place to only save work you've done with public data, like we've done here. To save your work locally, you will need Tableau Desktop, which costs approximately $900 a year.

TAKE THE SURVEY

Now that you're all done with the workshop, we'd so appreciate your feedback so that we can continue to improve our workshops for future attendees. Just click here to answer a few quick questions. Thank you!