ISSN: 2473-4901
v9 n5910

Teaching Case

Finding the Popularity of My Name over Time: Big Data Visualization

Frank Lee
flee@gsu.edu

Alexander Algarra
aalgarra1@student.gsu.edu

J. Mack Robinson College of Bueinss
Georgia State University
Atlanta, Georgia 30303, USA

Abstract

Exploratory data analysis (EDA), data visualization, and visual analytics are essential for understanding and analyzing complex datasets. In this project, we explored these techniques and their applications in data analytics. The case discusses Tableau, a powerful data visualization tool, and Google BigQuery, a cloud-based data warehouse that enables users to store, query, and analyze large datasets. It also explored the benefits and applications of both tools and their integration with other platforms and services. The project offers an introductory insight into Tableau's functionalities, employing a data file from the US Census Bureau via Google BigQuery.

Keywords: Big Data, Visualization, BigQuery, Tableau

1. INTRODUCTION

Overview

Data visualization is an essential tool in data analytics and is used across various fields and industries to communicate insights and findings from data. Visual elements like charts, graphs, and maps represent data and information graphically in data visualization (Healy, 2018). The purpose of data visualization is to simplify complex data so it can be understood and easily interpreted. It can assist in identifying patterns, relationships, and trends in large amounts of data, making them more accessible and actionable. The best data visualizations incorporate design and communication principles to make them aesthetically pleasing, easy to read and effectively convey the intended message (Healy, 2018).

Visual analytics uses interactive and visual methods to analyze and understand complex data. With visual analytics, you can explore and understand large and complex data sets flexibly and interactively by combining data mining, statistical analysis, and information visualization techniques. Visual analytics aims to enable decision-makers to gain insight and understanding of complex data sets rapidly and to support them in making informed decisions. Using visual analytics tools, users can explore and identify patterns in the data that may have yet to be apparent through other analysis methods (Keim, Andrienko, Fekete, Görg, Kohlhammer, & Melançon, 2010).

Throughout the project, it is crucial to consider the limitations of the data, including missing values, data quality issues, and biases, and to take appropriate steps to address these issues. Communication and collaboration with stakeholders, including domain experts, are essential to ensure that the insights and recommendations are relevant and actionable.

Tableau

Tableau is a powerful data visualization tool that allows users to create interactive and dynamic visualizations, dashboards, and reports. It is used for analyzing and presenting complex data simply and intuitively, making it easier for users to identify patterns, trends, and insights (Murray, 2018).

Google BigQuery

Google BigQuery is a cloud-based data warehouse that enables users to store, query, and analyze large datasets. It is part of the Google Cloud Platform and offers a highly scalable and cost-effective solution for managing big data (Valliappa Lakshmanan & Tigani, 2017). With BigQuery, users can quickly load and query massive amounts of data from various sources, including Google Cloud Storage, Google Drive, and third-party sources such as Amazon S3. The platform uses a columnar data storage format for faster query speeds and more efficient data processing.

Tableau and Google BigQuery Combined

The BigQuery public datasets can be used with Tableau to create powerful data visualizations to help users gain insights and make data-driven decisions (Valliappa Lakshmanan & Tigani, 2017). To get started, users can connect Tableau to their BigQuery account and access the public datasets available on the platform. Once the dataset is selected, users can create interactive visualizations, dashboards, and reports using Tableau's drag-and-drop interface and data visualization tools.

Learning Objectives

By completing this assignment, you will be able to:

- Apply the fundamental concepts of data visualization to define a project in your field of study.
- Practice the core principles using widely available tools (e.g., Tableau).
- Demonstrate the best practice that presents your story in creating data visualization, including connecting to different data sources, assessing the data quality, and converting raw data into data visualizations that provide actionable information.

2. CASE BACKGROUND

Case Text

Alex Smith has decided to leverage the power of data analytics techniques to answer the question: how common is the name *Alex* within the United States compared to other names starting with A? After recently taking a data analytics class at the university, *Alex* learned of the ability to access Big Data (BigQuery) public datasets and how to analyze the data using a visualization tool (Tableau), which has proven to be an invaluable resource in answering this question. Based on his data analytics experience, *Alex* understands that finding how common his name is can be extended to any name, and the process used can give others the option of learning the same.

Alex began by accessing the Social Security Administration's dataset, including all names from Social Security card applications for births in the United States after 1879. He imports the Google BigQuery dataset into Tableau, where he can easily query and analyze the data. Using Tableau, he created filters only to include names starting with the letter A and then grouped the data by name to count the number of occurrences for each name. With this information, *Alex* creates a visualization using Tableau that shows the top 10 most common names starting with A and the number of occurrences for each name. He can see that the name Andrew is the most common, followed by Anthony and Arthur.

Wanting to dig deeper, *Alex* creates a map visualization of the United States with bubbles representing the number of people named *Alex* in each state. The size of each bubble corresponds to the number of people with the name *Alex* in that state. By connecting Tableau to the BigQuery dataset, he can easily map the number of *Alex's* in each state.

Upon examining the map, *Alex* notices that the states with the highest number of people named *Alex* are California, Texas, and New York. However, *Alex* is curious to see if any regional trends or outliers may indicate other states where the name *Alex* is prevalent. Using Tableau's visualization tools, Alex breaks down the data by region and discovers that the Northeast and West Coast have the highest concentrations of people named *Alex*. He also notices that the name *Alex* is less common in the Midwest and the South.
After completing this project, *Alex* shared the step-by-step guide and the code used to analyze the data with others interested in finding out how common their names are. Other people can follow the same steps but with their name of interest

and use the same visualization tools to create maps and bar charts showing their name's frequency across different states and years.

To apply this project to other names, the user would need to modify the filters to include their name. They can then follow the steps outlined and use Tableau to create a map visualization or bar chart that shows the frequency of their name across different states and years. Overall, this project provides a framework and a set of tools that can be used by anyone interested in exploring the frequency of their name in the United States. The step-by-step guide can help other users replicate the process and apply it to their own names. By leveraging the power of big data and visualization tools, anyone can now explore the frequency of their name and gain insights into the popularity of their name across different states and years.

The Data Source

The project uses the public dataset created by the Social Security Administration hosted in Google BigQuery. The dataset contains all names from Social Security card applications for births that occurred in the United States after 1879. Note that many people born before 1937 never applied for a Social Security card, so their names are not included in this data. For others who did apply, records may not show the place of birth, and again their names are not included in the data. All data are from a 100% sample of records on Social Security card applications as of the end of February 2015. Once you load the dataset into Tableau, you will begin filtering your name by the total number of occurrences and geographical location.

3. PROJECT ACTIVITY

The purpose of this exercise is to become familiar with the process of analyzing a Google BigQuery dataset using the data visualization application Tableau. We will use the "USA Names" public dataset created by the Social Security Administration, which contains all names from Social Security card applications for births that occurred in the United States after 1879. The dataset will be used to explore how common your name is based on the given data.

The data was collected by the US Census Bureau. The instructions describe using some of the basic features of Tableau. After Connecting Your BigQuery Account to Tableau, please make sure to follow each of the steps in order.

Creating a Project and Connecting to a Data Source

1. Select "Project" under the project tab and type "bigquery-public-data" (Appendix A, Figure 1). Once you have completed this step, you can access the "USA Names" dataset in the "bigquery-public-data" project and begin analyzing it in Tableau.
2. After selecting the "bigquery-public-data" project, the next step is to choose the specific dataset you want to analyze (Appendix A, Figure 1). In this case, we are using the "USA Names" public dataset the Social Security Administration created. In the "Select a dataset" field, click on the drop-down arrow and select "usa_names." By selecting the "usa_names" dataset, Tableau will load the data from this dataset into the visualization environment, allowing you to create visualizations and perform data analysis on this specific dataset.
3. After loading the data from the "usa_names" dataset in Tableau, you will be directed to the worksheet area where you will see a list of available sheets. Click on the "Sheet 1" tab to create a new sheet where you can start building your visualization.

Information Filtering

1. Once you have added the data source to the report, the next step is to create a visualization. First, let's look at the most common names beginning with the letter "A." The first step is to create a filter for the "Name" variable, allowing you to narrow down the dataset to show only the names that start with the letter "A".

2. When working with data in Tableau, filters are a powerful tool that can be used to narrow down the data to focus on specific subsets of interest (Appendix A, Figure 2). To create a filter in Tableau, you can begin by selecting the "Add" button within the data source tab.

3. Next choose "Add" within the edit data source filters. After selecting the "Add" button in the Data Source tab, you will be directed to the "Edit Data Source Filters" dialog box. In this dialog box, you can create a filter by selecting the "Add" button. Once you click on "Add," you will see a list of all the variables in your data source. In this case, you will see a list of all the variables in the "USA Names" dataset.

4. Select the "Name" variable from this list, as this is the variable we want to filter on. Once you have selected the "Name" variable, you will see several filter options, including the

ability to filter by individual values, ranges of values, or to use a wildcard filter.

5. Click the "Wildcard" option in the filter dialog box (Appendix A, Figure 3). This option will open a new dialog box where you can select the type of wildcard you want to use. In this case, you want to filter the names that start with a specific letter, so you should choose the "Starts With" category. Then, you can insert the first letter of your first name in the text field provided. For example, if your name is *Alex*, you would insert the letter "A" in the text field. It will filter the dataset only to show names that start with the letter "A".

6. After creating the filter for the "Name" variable in Tableau, the next step is to create a filter for the "Gender" variable. To do this, repeat step 5 for the "Gender" variable and select your gender. After following these steps, the dataset will be filtered only to include names matching the first letter of your name and gender.

7. After adding a filter in the data source tab or importing new data, you need to update the data set to ensure that it reflects the changes made. To do this, click on the "Update Now" button which is located at the bottom of the data source tab in Tableau. It will refresh the data and apply any new filters or changes that have been made. It is important to note that failing to update the data set can result in inaccurate or incomplete visualizations, so it is always important to double-check that the data is up to date before proceeding with any analysis or visualization.

Creating Two Basic Charts

1. Once in the Sheet view, locate the "Measures" section is found in the Data pane on the left-hand side of the screen (Appendix A, Figure 4). Find the measure called "usa_1910_current" and click and drag it over to the "Columns" shelf located on the top right-hand side of the screen. Next, find the "Dimensions" pane located below the "Measures" pane. Click and drag the "Name" variable over to the "Rows" shelf on the screen's right-hand side. Once both the "usa_1910_current" measure and "Name" variable have been added, you should see a table of data displayed in the main view of the Sheet.

2. Because the Name category contains many names, we will add a filter for "Name" within Sheet 1 (Appendix A, Figure 5). To start, please click and drag the name category to the Filters section, and the Filter dialog box will appear.

3. Once in the filter, choose the "Top" tab, which allows you to filter the data based on the top values in a particular field (Appendix A, Figure 6). Once you have selected the "Top" tab, you can choose the "By field" option to select a field to determine the top values and then choose the "Top 10."

4. Choose "Sort Descending" to sort the names in descending order based on the count (Appendix A, Figure 7). You should now see a bar chart that displays the top 10 names starting with the first letter of your name, sorted by the number of occurrences in descending order. In the graph shown below, we have sorted the names descending by count and you quickly see that *Alex* is the 9th most common name in the dataset.To make the graph more visually appealing, we can edit the color palette by dragging "Name" and dropping it in the Color tab under Marks (Appendix A, Figure 8). Once you have done this, Tableau will use the Automatic color palette by default. If you want to change the colors, click on the Name legend on the right side, as shown below. You can choose from a variety of pre-built color palettes or create your own custom color palette by selecting "Palette" from the drop-down menu. To create a custom color palette, you can select individual colors by clicking on the color wheel or choosing a color scheme from the drop-down menu.

5. Once you have applied the filter and edited the color palette, the completed graph should be displayed in the "Sheet1" tab. The graph will display the top 10 names that start with the first letter of your name, based on the number of occurrences in the dataset. The bars on the graph represent the count of occurrences for each name, and the color of the bars corresponds to the color palette chosen. The x-axis displays the names, and the y-axis displays the count of occurrences. The final graph should look similar to the example provided in the appendix.

6. Tableau offers numerous types of charts that are great for data visualization. For example, Tableau can format the data into the following:
   - Bar charts: A common type of chart used to represent numerical data using rectangular bars, where the length of

each bar represents the value of the data being displayed.
- Line charts: Used to show trends in data over time, with data points connected by a line.
- Pie charts: Circular graphs that are divided into slices to represent the relative sizes of different categories of data.
- Maps: Used to visualize geographic data and display information based on location.
- Density maps: Use color to indicate the concentration of data points in specific areas.
- Scatter plots: Used to visualize the relationship between two variables and can help identify patterns in data.
- Gantt charts: Commonly used in project management to show the duration of tasks and their dependencies.
- Bubble charts: Like scatter plots, but the size of the marker represents the value of a third variable.
- Tree maps: Use nested rectangles to display hierarchical data, with larger rectangles representing higher-level categories and smaller rectangles representing subcategories.

7. For example, we will create a Bubble Chart, in a new sheet (Appendix A, Figure 9). First right click the Sheet 1 tab and rename the sheet to “Bar Chart” and then select new sheet and rename Sheet 2 as “Bubble Chart”. After renaming the Sheet 1 tab to "Bar Chart" and creating a new sheet named "Bubble Chart," we can start building our Bubble Chart by selecting the desired variables. Once you have opened the new sheet we have created, to start, click and drag usa_1910_current (Count) to the Color and Size Tab in the Marks section. Click and drag Name to the Label Tab in the Marks section. After completing these steps, Tableau automatically analyzes what you are attempting to accomplish and automatically generates a bubble chart for you based on the specified criteria.

Creating a Map Chart

We now have two good charts for this data visualization, and it is time to move on to find how many people in each state share your name. To create a visualization that shows how many people in each state share your name, we need to create a new sheet. First, right-click on any existing sheet, select "New Worksheet," and name it “Map.”

1. Drag the "Name" variable to the "Filters" shelf, just like we did earlier in the Bar Chart sheet (Appendix A, Figure 10). In the "Filter Field" dialog box, choose the "Wildcard" option, select “Exactly matches,” and enter your name in the “Match Value area.” After that, click on the "Apply" button to filter the data by name.

2. Now that we have applied the filter, we can begin specifying the variables to its appropriate categories (Appendix A, Figure 11). First, click and drag the "latitude" variable to the "Rows" shelf. The "latitude" variable represents the geographic coordinate that specifies the north-south position of a point on the Earth's surface. By dragging "latitude" to the "Rows" shelf, we are telling Tableau to use this variable to define the vertical axis of the map. After dragging the latitude field to the Rows section, the next step is to click and drag the longitude field to the Columns section.

3. Now, we will add the size and color of the data points. Here, we want to use the count of occurrences of the selected name as the size and color of the data points. To do this, click and drag the usa_1910_current variable to the "Size" and "Color" options in the "Marks" section. The "Size" option controls the size of each data point on the map, while the "Color" option specifies the color of each data point based on a chosen variable. In this case, we want to use the count of occurrences of the selected name for both the size and color so that we will drag and drop the same variable to both options. After dragging and dropping, you will see the map update with the selected name's count of occurrences in each state reflected in the data points' size and color. Finally, click and drag the "State" field from the "Data" pane to the Detail area on the Marks card. It tells Tableau to group the marks by state, so you can see the number of people with your name in each state.

4. Tableau has automatically created your Map. You may add a filter for the State measure to remove Alaska and Hawaii. You may also change the background of the map to make it more visually appealing.

5. Now we have created a Map visualization. Let’s look at how visualizing this data looks like in a bar chart (Appendix A, Figure 12). In this case, we want to create a bar chart to show the top 20 states with the most names. We will need to use the same data source as

our previous visualizations to do this. First, create a new sheet and rename it “Top 20 States with the Most Names”. The next step is to click and drag the usa_1910_current measure to Columns and click and drag State to Rows and Colors. It will create a vertical bar chart where the height of the bar corresponds to the number of people with the given name in each state. In this chart, the x-axis represents the different states, and the y-axis represents the number of people with the given name in each state. The bars are colored according to the color palette selected, and you can hover over each bar to see the exact number of people in that state with the given name.

6. To create a filter for the Name measure, first click on the "Name" dimension in the "Data" pane on the left side of the screen. Then, drag the "Name" dimension to the "Filters" shelf at the bottom of the screen. This will open the "Filter" dialog box and choose the Wildcard option only to include your name.

7. To create a filter for the State measure and choose the top 20 states first click on the "State" dimension in the "Data" pane on the left side of the screen. Then, drag the "State" dimension to the "Filters" shelf at the bottom of the screen. In the "Filter" dialog box, select the "Top" tab. In the "Top" tab, select "By Field" and choose "usa_1910_current" from the drop-down menu of available fields. Set the "Top" filter to "20" (or any other desired number). Click "OK" to apply the filter. Next, change the color palette to one you think is the most effective.

Dashboard and Export

1. Finally, to create a more shareable output, we export the visualizations to PowerPoint. Tip: To optimize a dashboard for PowerPoint, on the Dashboard tab, choose Size > Fixed Size > PowerPoint (1600 x 900). To begin, create a new Dashboard tab and rename it to Your_Name_Data_Visualization. You create a dashboard much like you create a new worksheet.

2. From the Sheets list left, drag views to your dashboard on the right. Now we can click and drag the four visualizations into the dashboard area (Appendix A, Figure 13).

3. To export your new dashboard to PowerPoint, you must first click File on the top right side of the screen, next click Export as PowerPoint… In the following prompt click Select All, and finally, click Export. Tableau will then ask you to save the file to your device. Please save it under a file name of your choice and open the PowerPoint (Appendix A, Figure 14).

4. Once you open the PowerPoint, you should see your visualizations in individual slides, including the title, bar chart, bubble chart, map, and the top 20 states bar chart. From here, you can begin formatting your PowerPoint and adding the stories to your slides to make the data exploration come to life (Appendix A, Figure 15).

4. PROJECT REPORT

The task in this assignment is to use Big Data (BigQuery) and a visualization tool (Tableau) to formulate and answer a (series of) specific question(s) about a data set of your choice and then write a story about the data. After answering the questions using the data, you need to create final visualizations that tell the story of the data. Finally, write a couple of paragraphs describing the story, the visualization, and how it answers your questions.

5. REFERENCES

Healy, K. (2018). Data Visualization: A Practical Introduction. Princeton University Press.

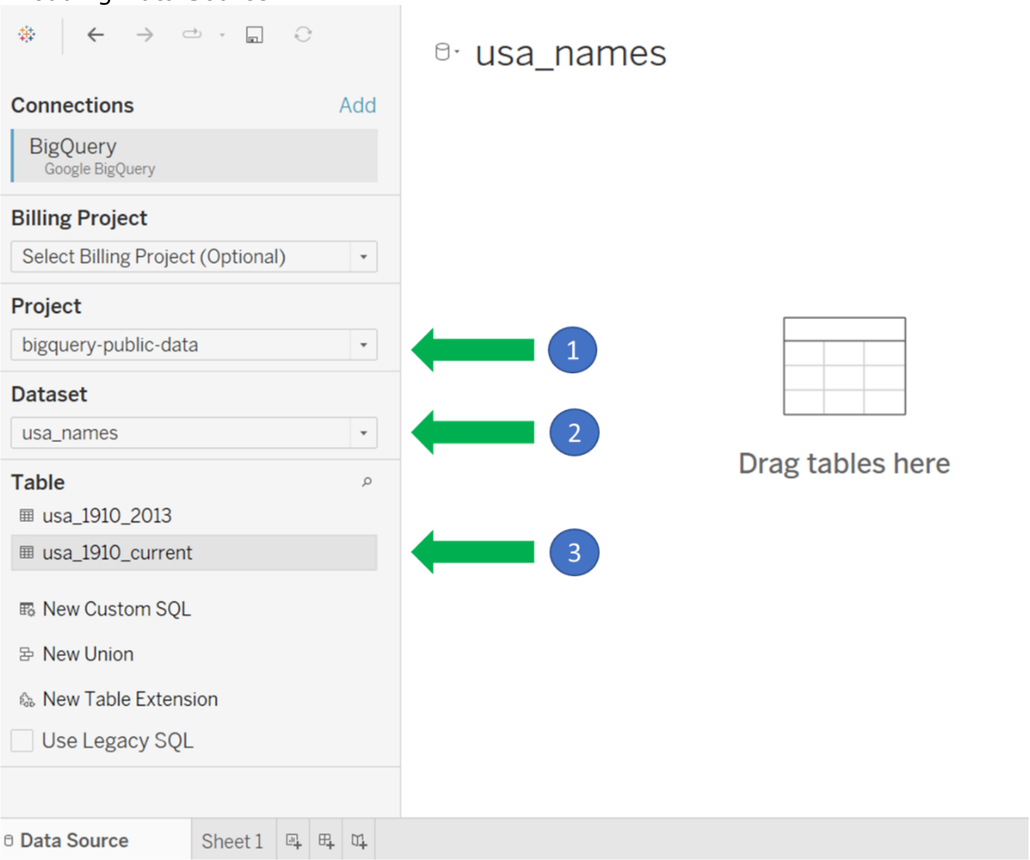
Keim, D., Andrienko, G., Fekete, J. D., Görg, C., Kohlhammer, J., & Melançon, G. (2010). Visual Analytics: Definition, Process, and Challenges. In Information Visualization (pp. 154-175). Springer, Berlin, Heidelberg.

Murray, R. (2018). Tableau Your Data!: Fast and Easy Visual Analysis with Tableau Software. Wiley.

Valliappa Lakshmanan, V., & Tigani, J. (2017). Google BigQuery: The Definitive Guide: Data Warehousing, Analytics, and Machine Learning at Scale. O'Reilly Media, Inc.

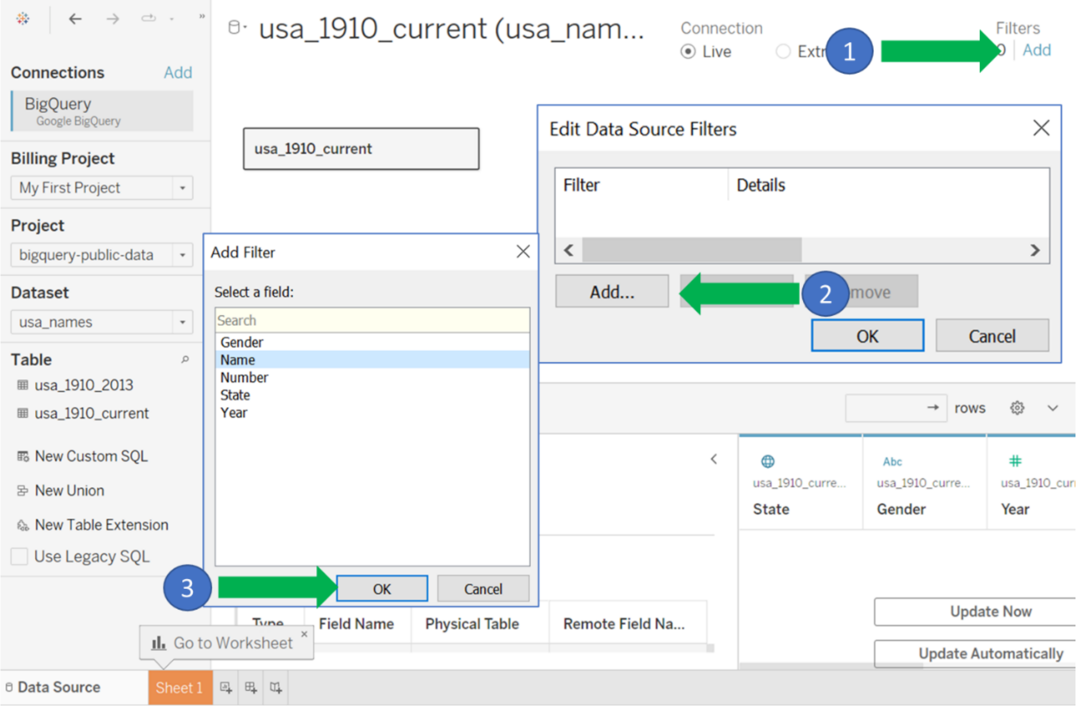
APPENDIX A
Guided Step-by-Step Figures

Figure 1. Loading Data Source



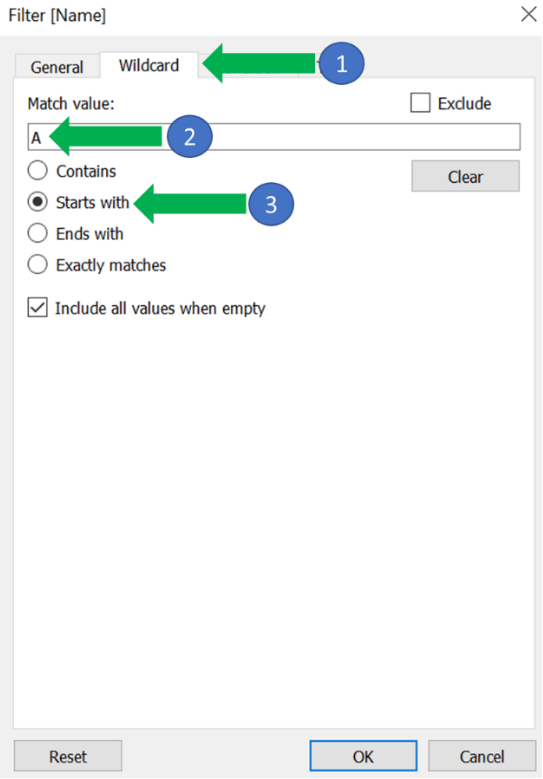
Connections: BigQuery (Google BigQuery)
Billing Project: Select Billing Project (Optional)
Project: bigquery-public-data
Dataset: usa_names
Table: usa_1910_2013, usa_1910_current

Figure 2. Filter by Variable



Billing Project: My First Project
Project: bigquery-public-data
Dataset: usa_names
Table: usa_1910_2013, usa_1910_current
Connection: ● Live ○ Extr
Add Filter – Select a field: Gender, Name, Number, State, Year

Figure 3. Filter by Wildcard



Filter [Name]

General | Wildcard

Match value: A

□ Exclude

○ Contains

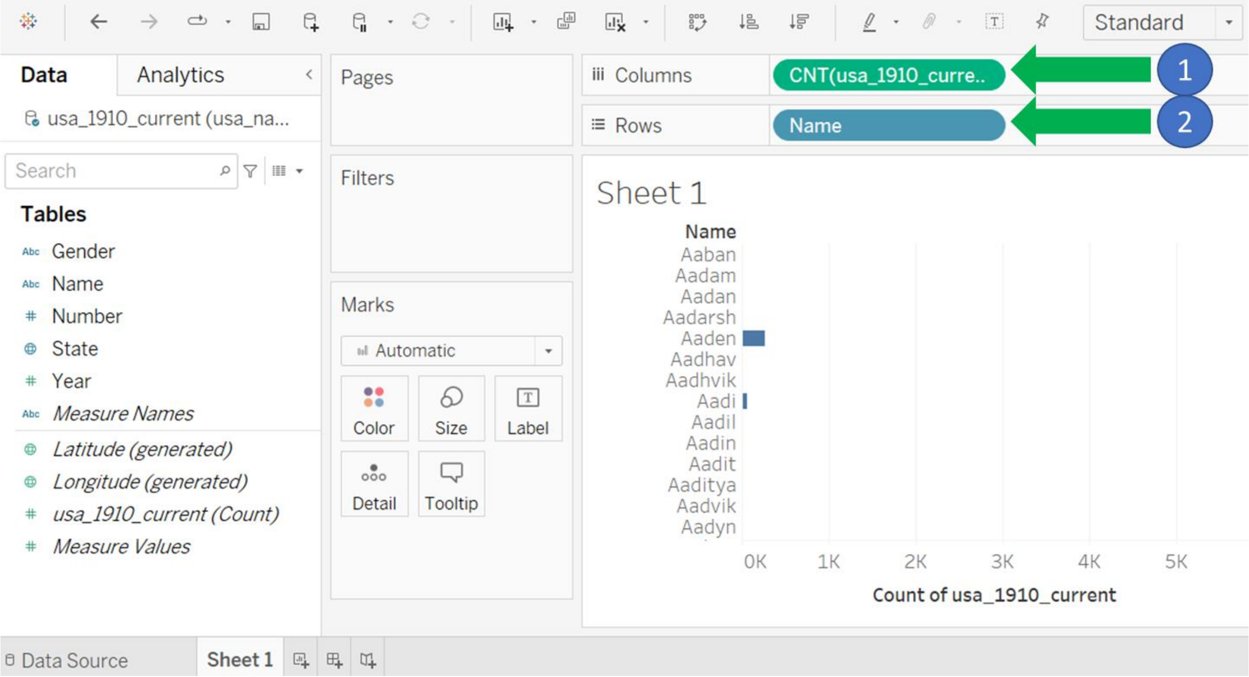
● Starts with

○ Ends with

○ Exactly matches

☑ Include all values when empty

Figure 4. Inserting Data into Columns and Rows



Columns: CNT(usa_1910_curre..

Rows: Name

Sheet 1

Name
Aaban
Aadam
Aadan
Aadarsh
Aaden
Aadhav
Aadhvik
Aadi
Aadil
Aadin
Aadit
Aaditya
Aadvik
Aadyn

Count of usa_1910_current

Figure 5. Applying Filter to Name

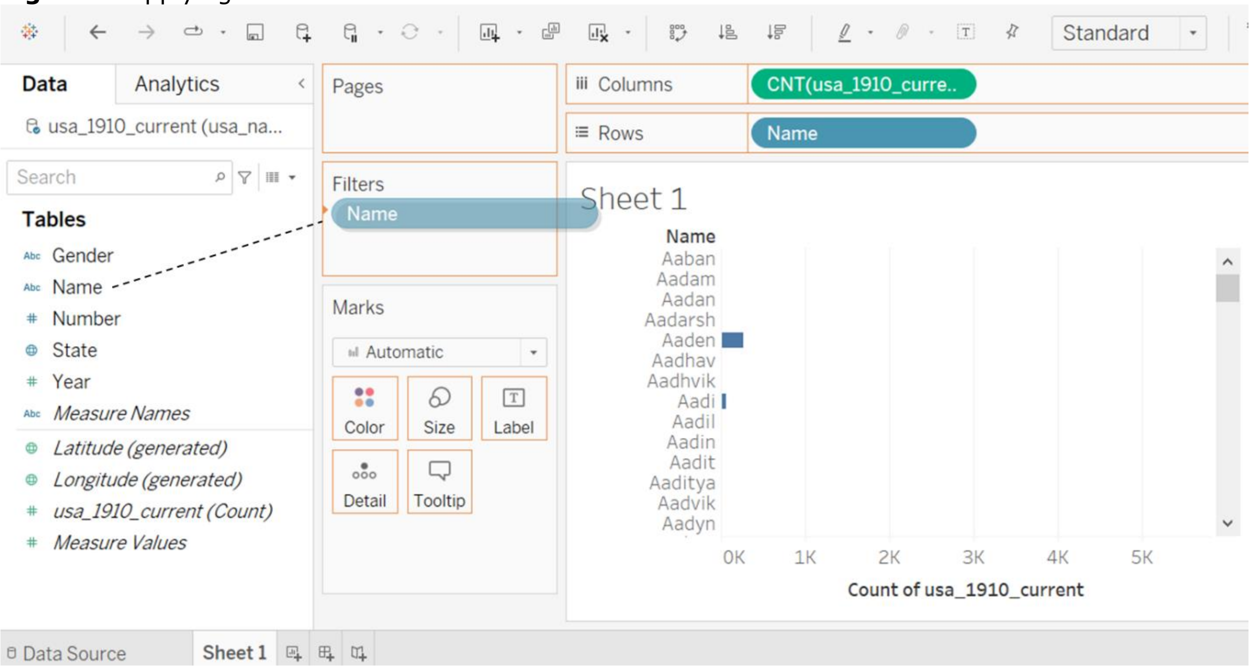
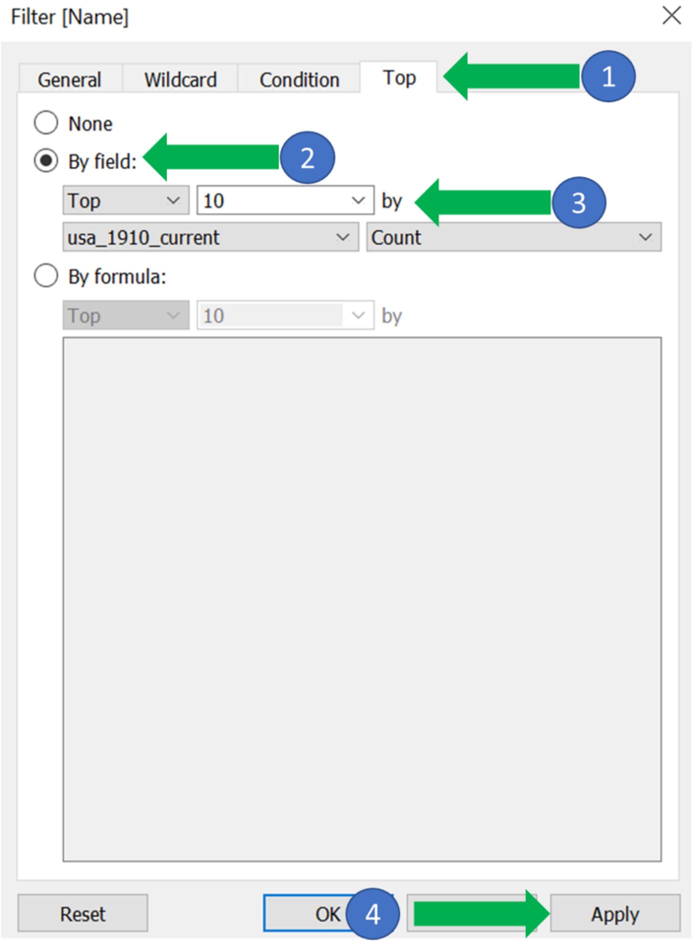
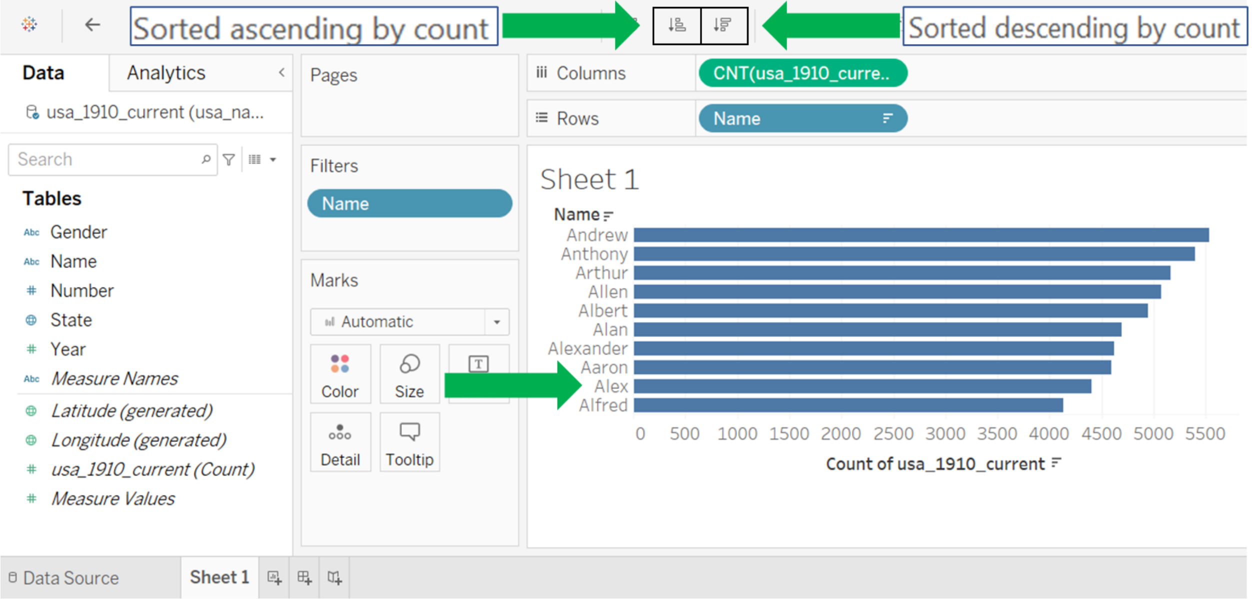


Figure 6. Top 10 names


Filter [Name]

○ None
● By field:
Top 10 by
usa_1910_current Count
○ By formula:
Top 10 by

Figure 7. Sort Ascending and Descending by Count.



Sorted ascending by count

Sorted descending by count

Figure 8. Edit Colors.

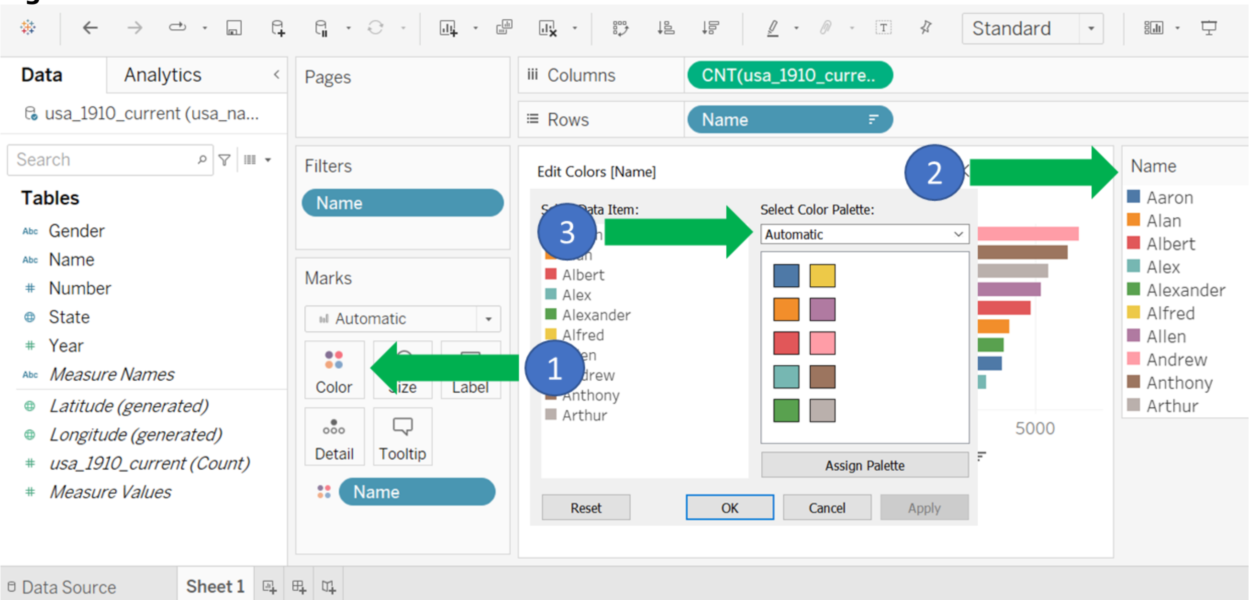


Figure 9. Bubble Chart

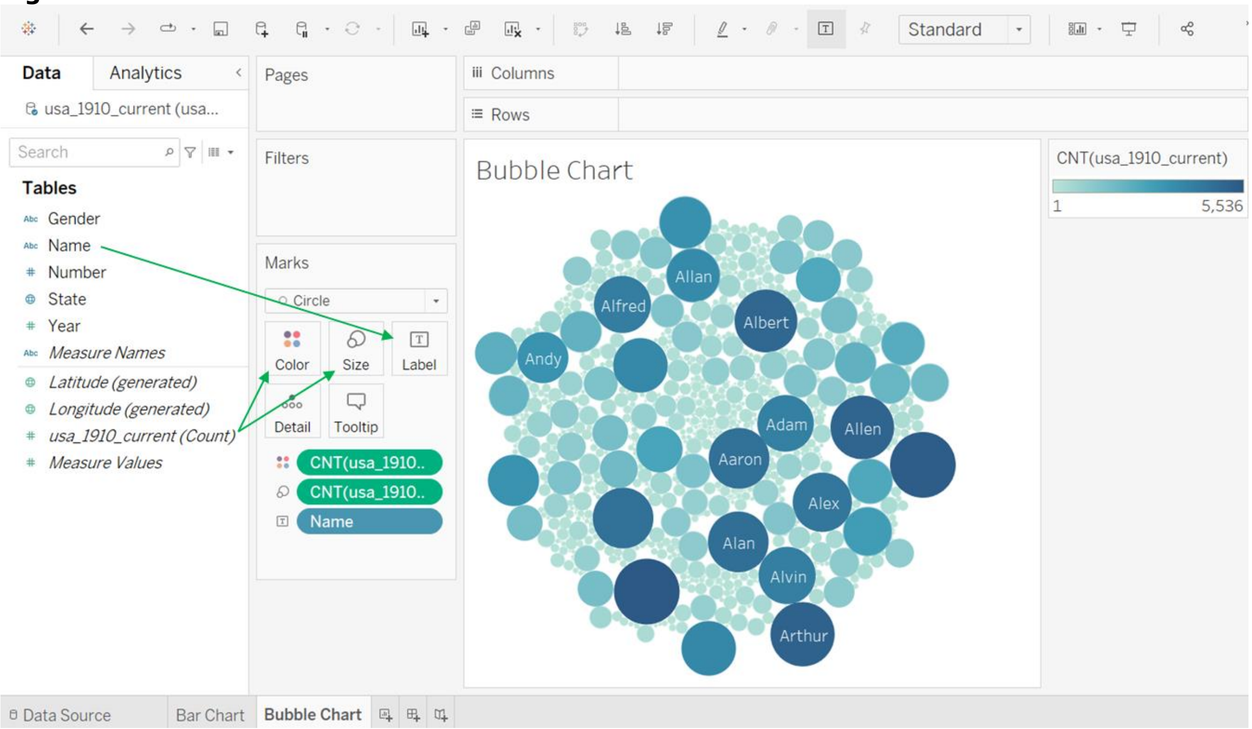
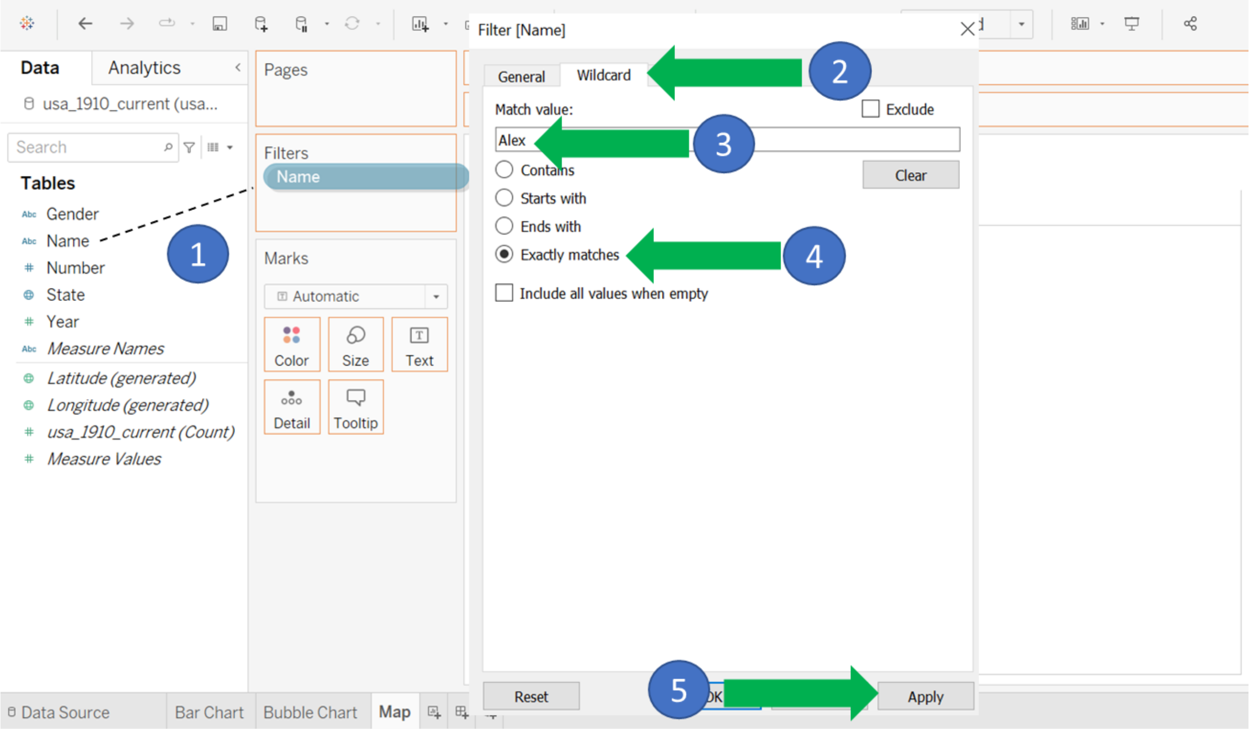


Figure 10. Name Filter for Map



Match value: Alex

□ Exclude

○ Contains
○ Starts with
○ Ends with
● Exactly matches

□ Include all values when empty

Figure 11. Creating a Map Chart

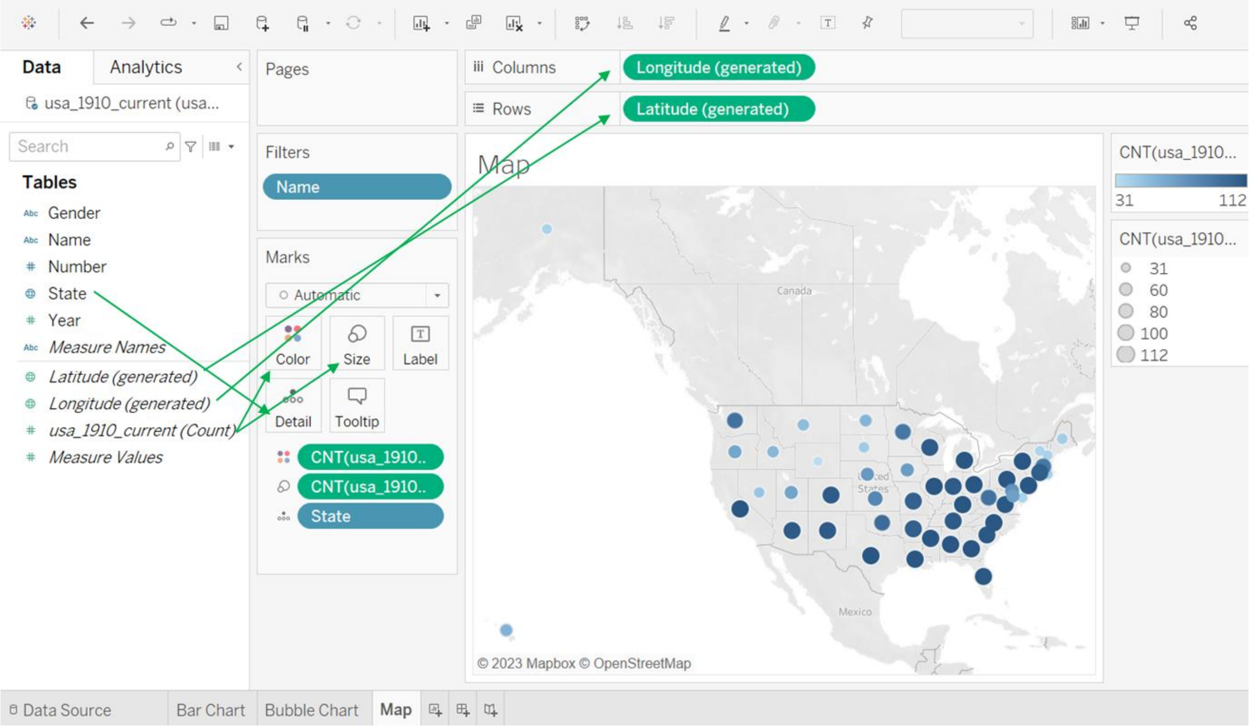


Figure 12. Bar Chart Containing Total Name Count by State

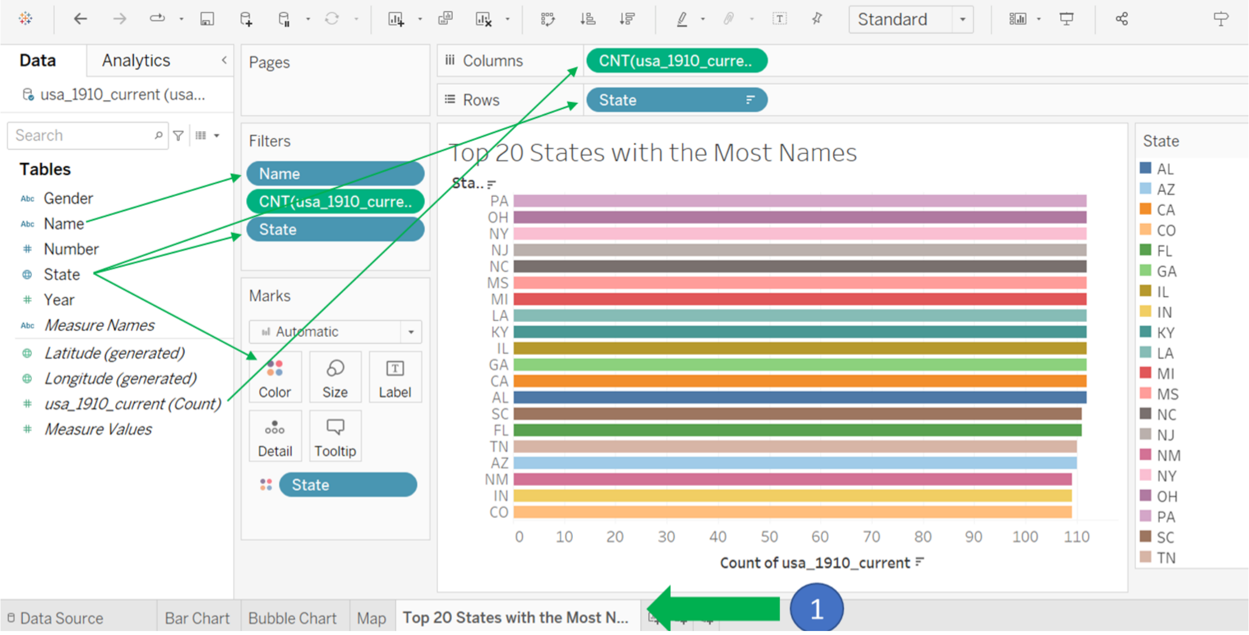


Figure 13. Visualizations Uploaded to Dashboard.

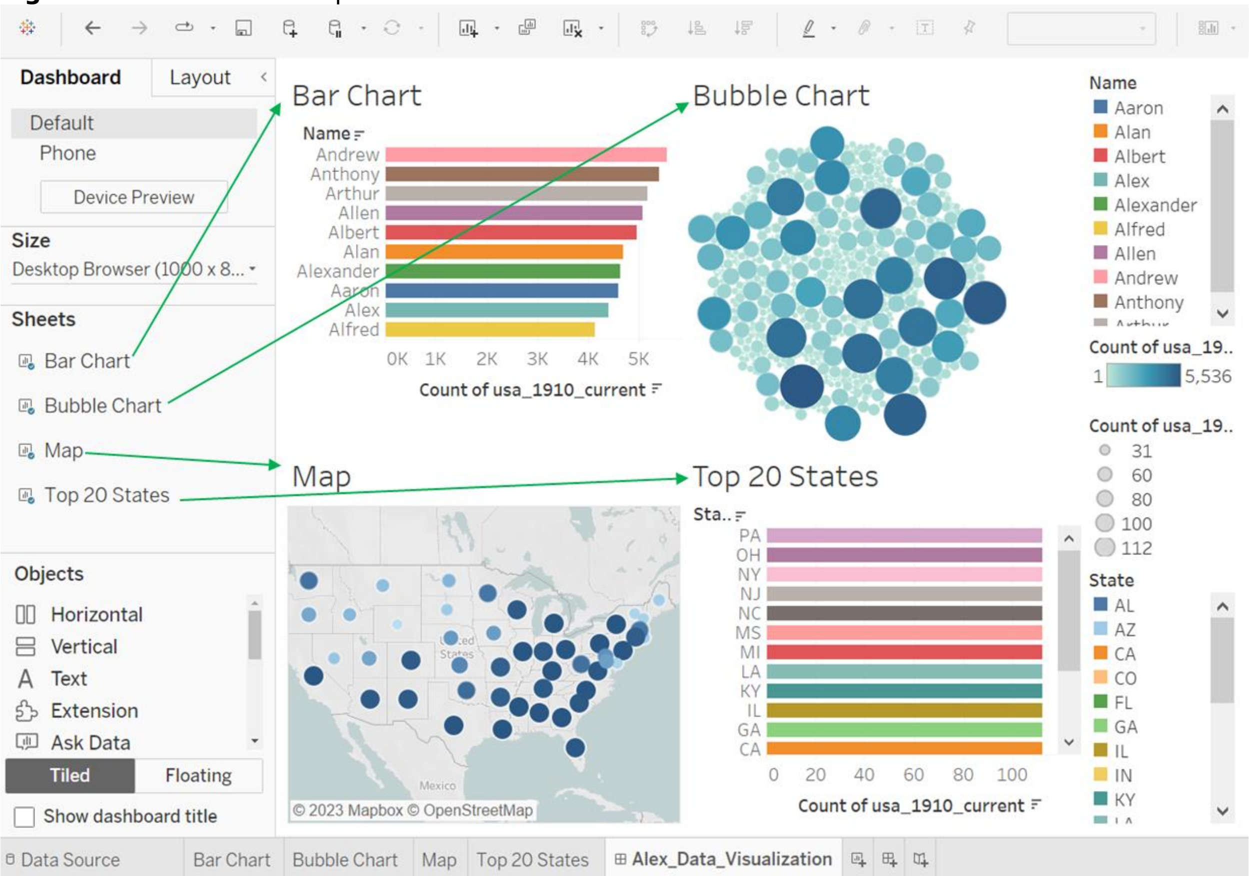


Figure 14. Export the New Dashboard

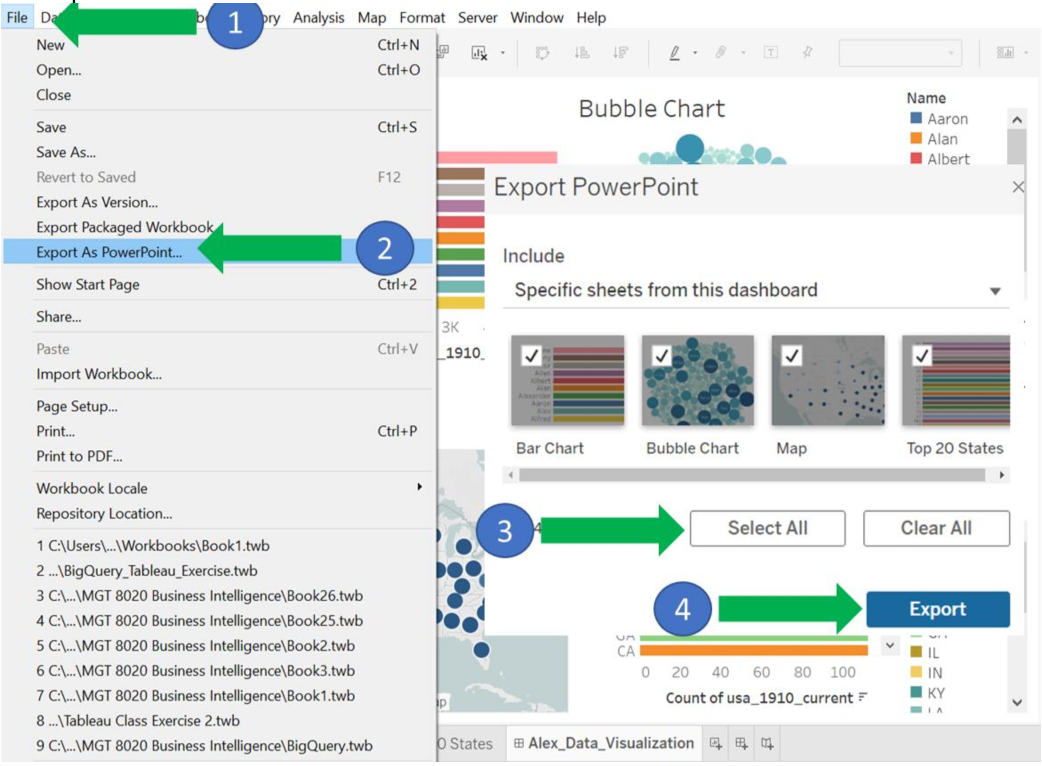
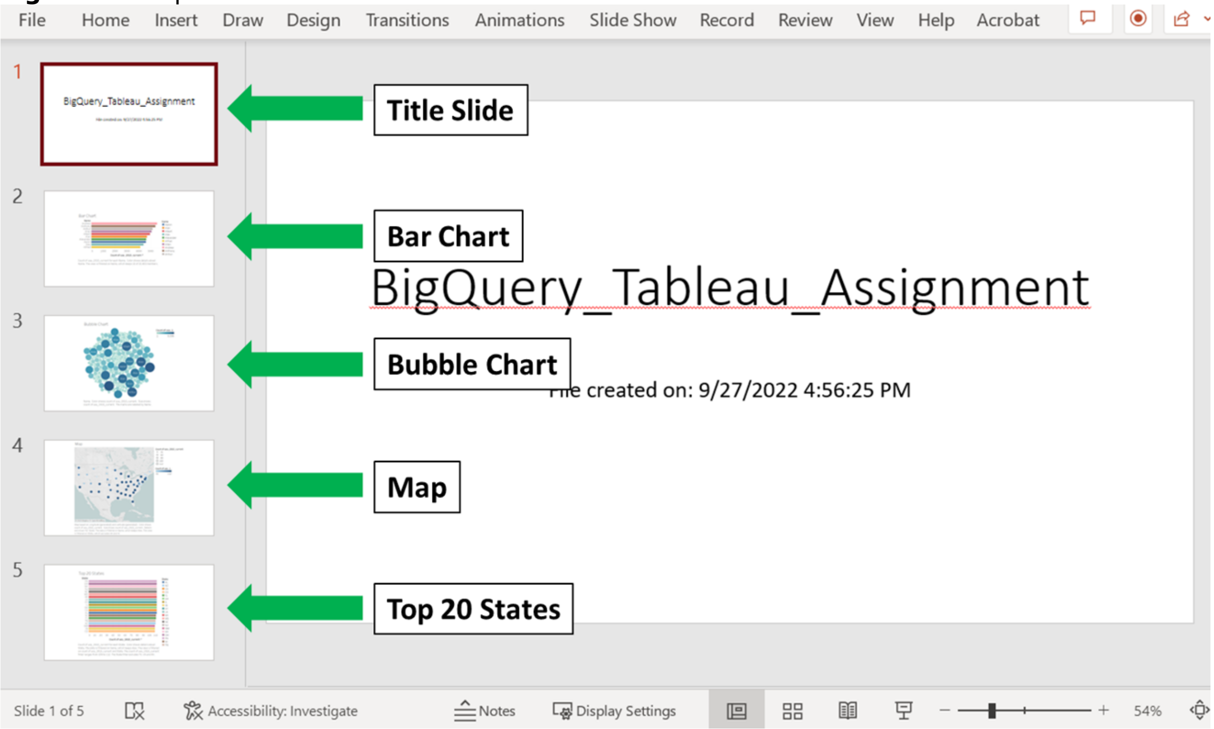


Figure 15. Exported PowerPoint



Title Slide

Bar Chart

Bubble Chart

Map

Top 20 States

BigQuery_Tableau_Assignment

File created on: 9/27/2022 4:56:25 PM