Data Visualization Tool: Tableau

CPS 563 – Data Visualization

Dr. Tam Nguyen

tamnguyen@udayton.edu

Introduction

- Tableau is a powerful data visualization software.
- Capable of creating various interactive visualizations from a multitude of data sources.
- Tableau is a commercial software, but is available to students for free.
  - Download from (http://www.tableau.com/academic/students)
- Tableau is primarily a **drag-and-drop** software.

Features

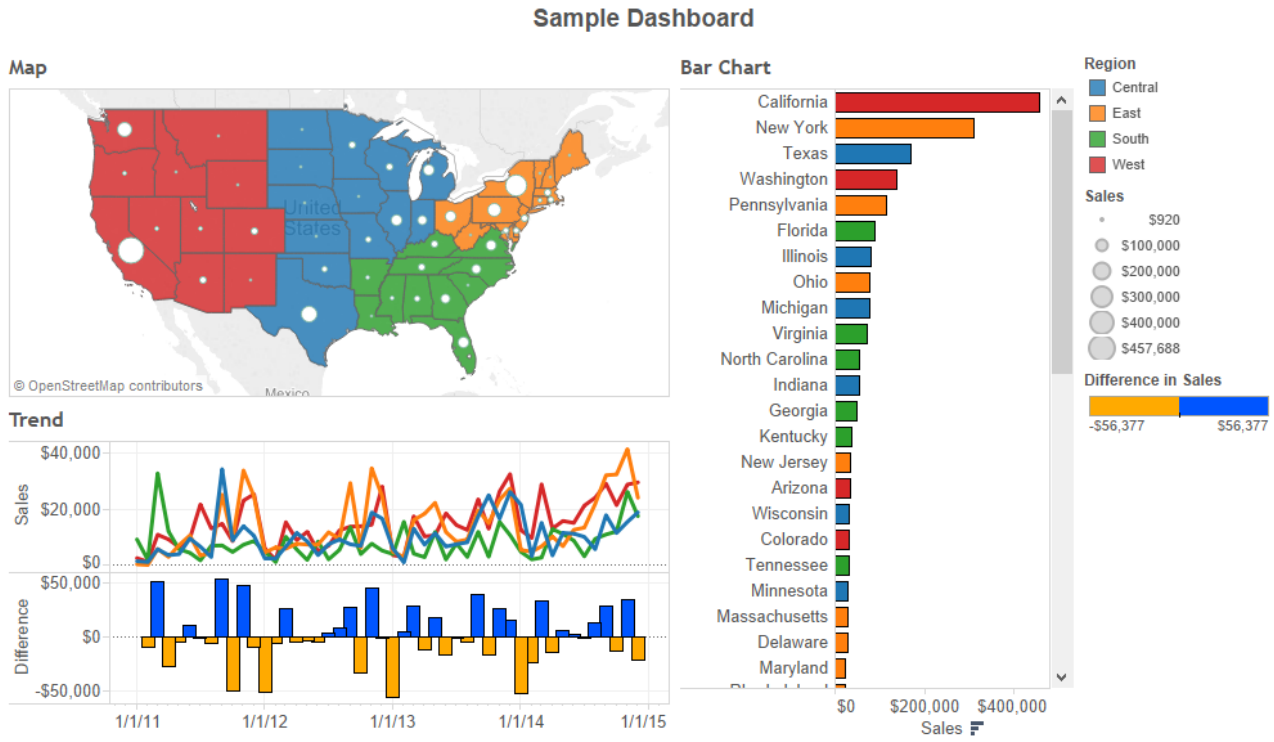
- Tableau can connect to variety of data sources, including:
  - Local files – Excel, Text, Access
  - Traditional databases – SQL Server, MySQL, Oracle, PostgreSQL, DB2
  - Big Data Technologies – Hadoop, Hive, Spark SQL

Features

- Tableau can create a variety of visualizations including:
  - Bar and line charts
  - Geospatial analysis
  - Word clouds
  - Treemaps

Features

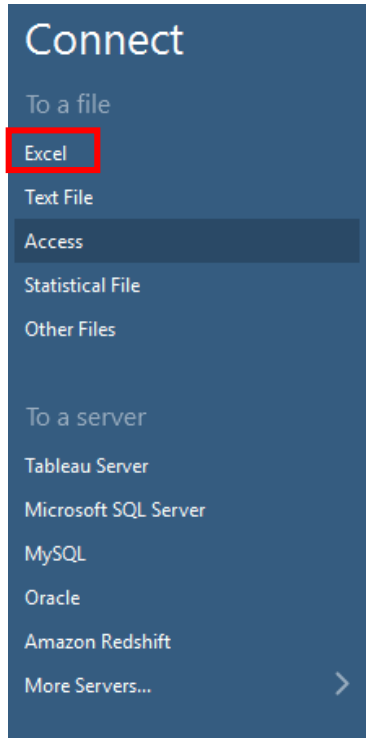
- These visualizations can be combined into interactive dashboards.
  - Can later be published online or shared easily.

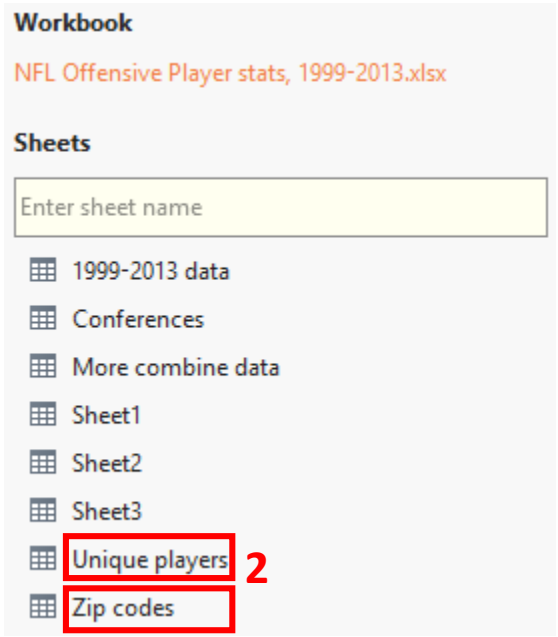


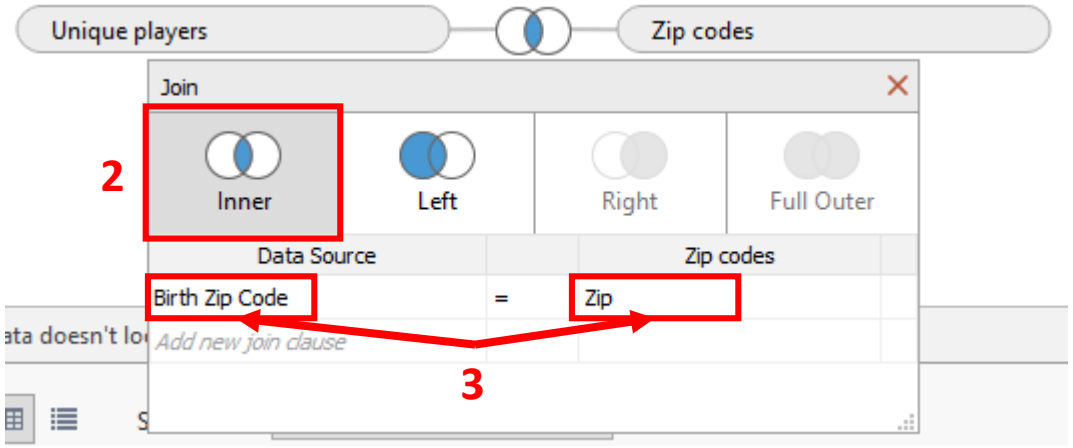
Examples

- The following examples show how to load data into Tableau, make three basic visualizations, and put them into a dashboard.
  - Bar chart, Word Cloud, and Geospatial visualization.
- Download "**NFL Offensive Player stats, 1999-2013.xlsx**" from isidore
- This is an Excel spreadsheet about **NFL Offensive players from 1999-2013**. It contains:
  - ~40,000 rows of data
  - Player information (physically measurable traits, birthplace, college attended)
  - Positions played
  - Wins achieved in career

Connecting to a Data Source





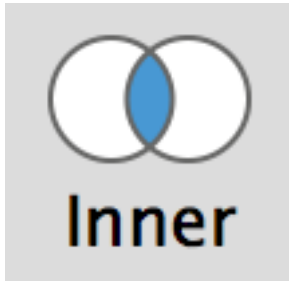


- We will have to connect to a data source to start making visualizations.
  1. Since our data is in an Excel workbook, we will select that.
  2. Second, we will join two of the sheets in the workbook such that we can get access to a larger set of data. Drag the “Unique players” and “Zip codes” sheets to the right. Select the “Inner” join option.
  3. We will join the sheets based on zip code.

Joins

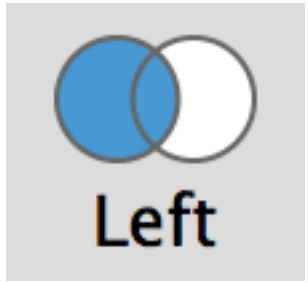
A	B
1	1
2	null
3	2
1	null
2	1

B	C
1	A
2	B
3	C



A	B	C
1	1	A
3	2	B
2	1	A

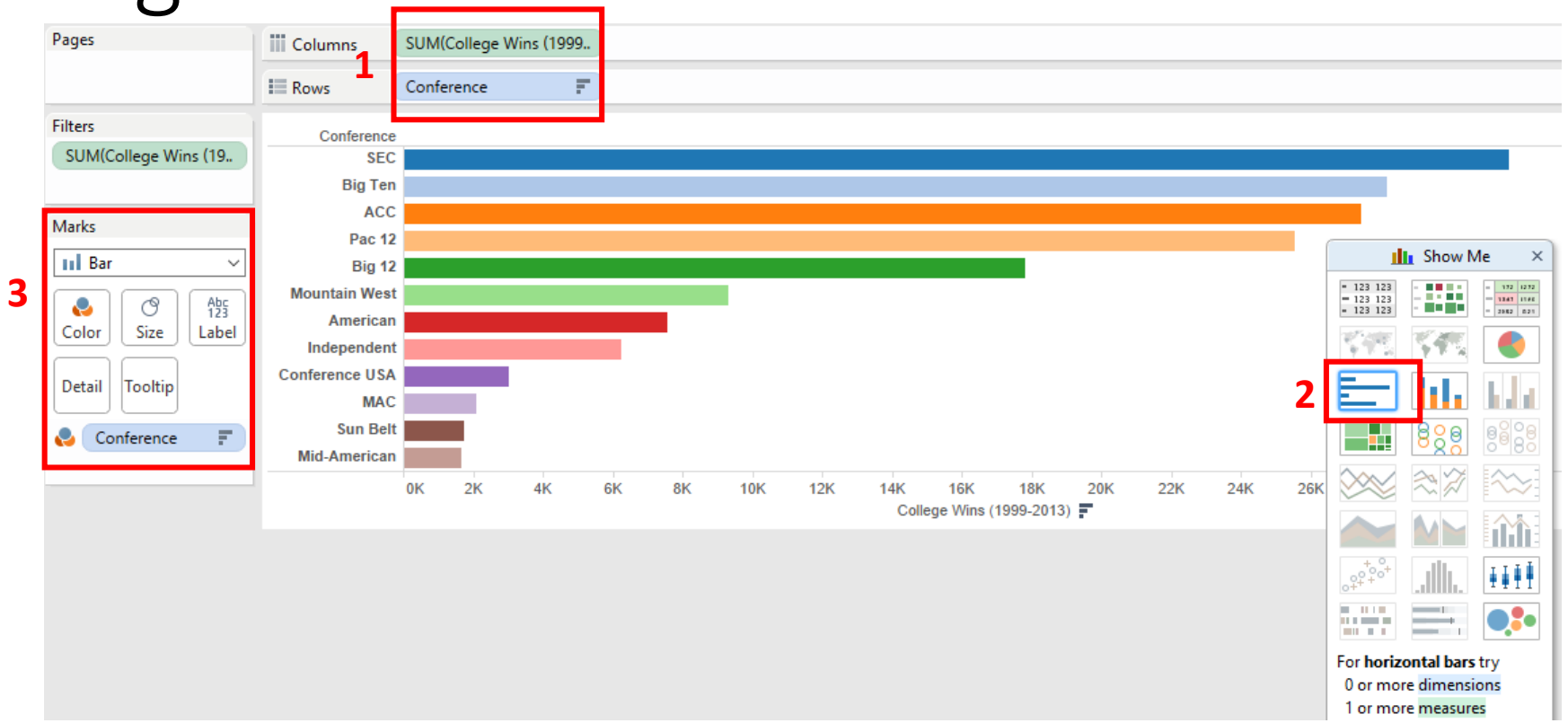
Only rows that have matches are included, so rows with no match are dropped.



A	B	C
1	1	A
2	null	null
3	2	B
1	null	null
2	1	A

Every row from the left table is included, unused rows from the right are dropped.

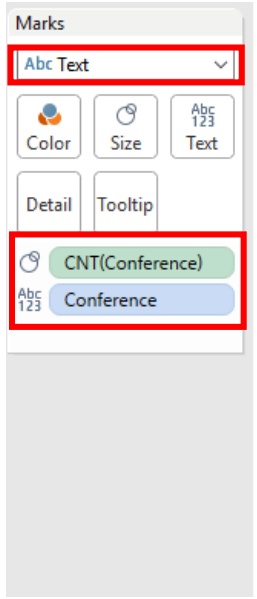
Creating a Bar Chart



- Suppose we want to know which major college conferences **have most combined wins since 1999**.

1. First, drag the “Conference” dimension into the “Rows” bar, and the “College Wins” into the columns. Hit the drop down on the “College Wins” and select “Sum.”
2. Second, select bar chart on the right hand side.
3. To add a little bit of color, drag the “Conference” into the “Color” mark.

Creating a Word Cloud



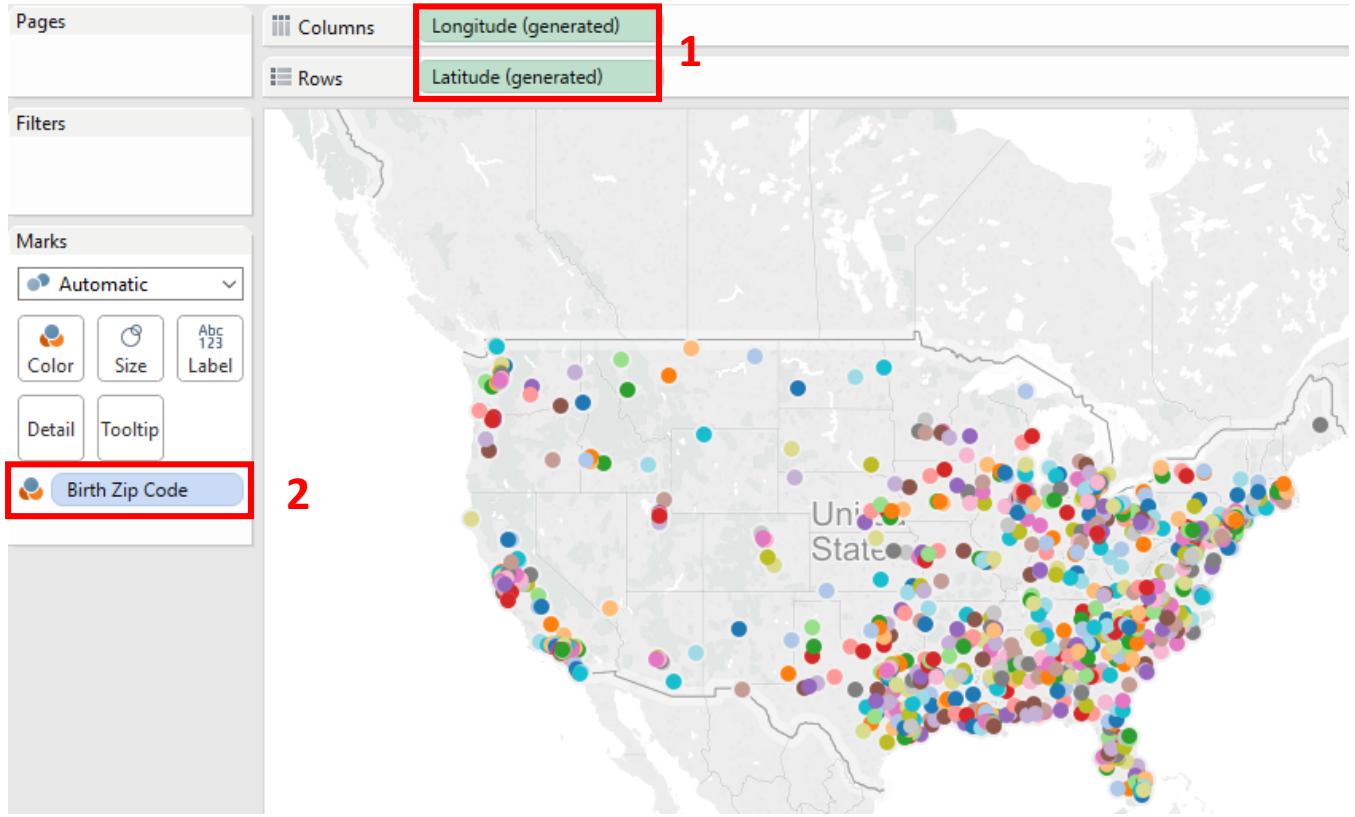
- Suppose now we want to get a general sense of the most popular conferences in terms of player enrollment is concerned. A word cloud is a great way to visually represent this.

1. First, switch the “Marks” option to “Text”.
2. Second, drag the “Conference” dimension into the “Text” marks box.
    1. Then drag the “Conference” dimension into the “Size” marks box.
    2. Adjust the measurement on this by hitting the drop down and selecting “Measure (Count)”

Creating a Geospatial Visualization

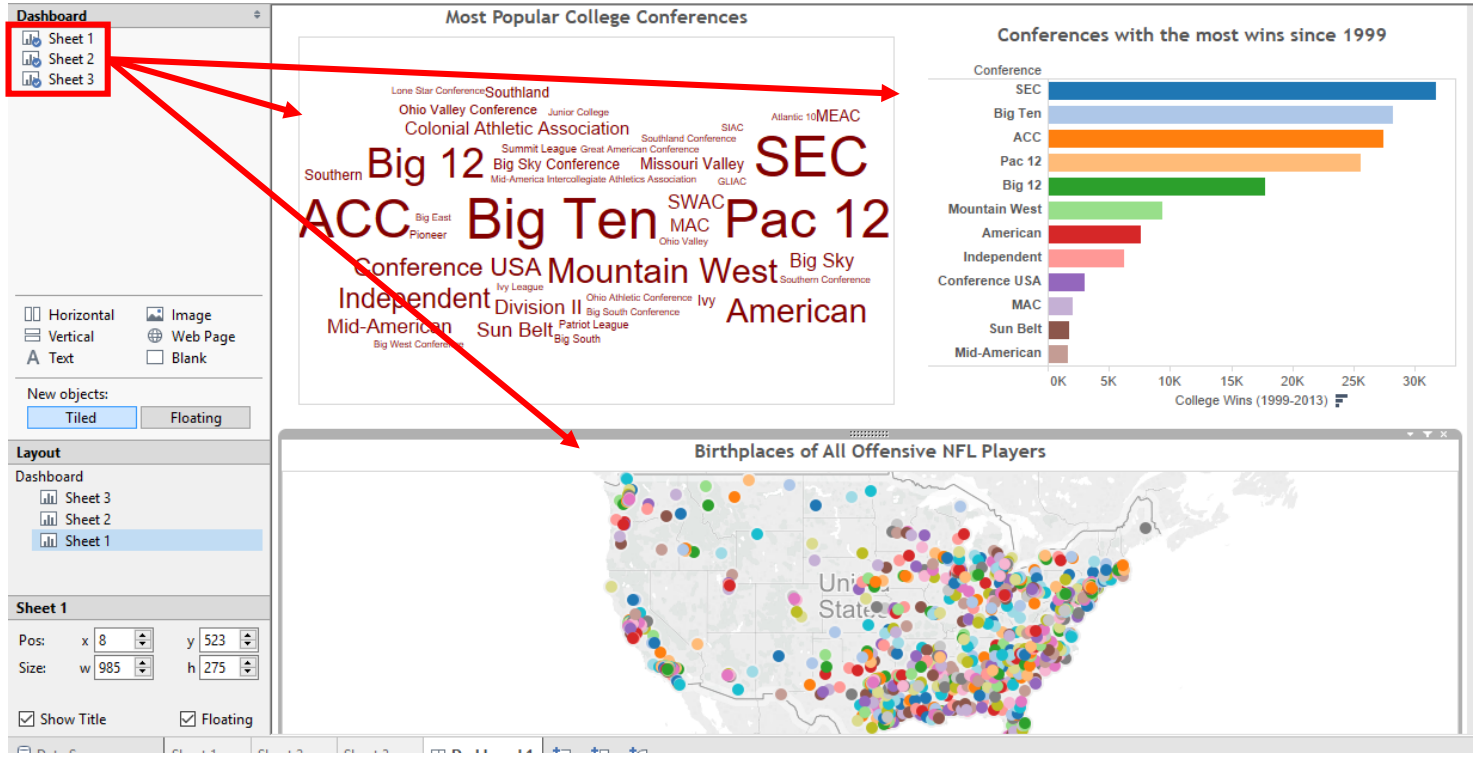
- Consider now that we are interested in the birthplaces of all of the NFL players.
- We can easily create a map representation.

1. Drag the “Longitude” dimension to columns, and “Latitude” dimension to the rows. Select the map visualization.
2. Add in some color by dragging the “Birth Zip Code” into the “Color” Marks.



Combining Visualizations into a Dashboard

- To tell a more comprehensive story, we can create a dashboard combining all of the visualizations.
- Simply open a dashboard view and start dragging sheets into the dashboard.
- You can format and add filters into the dashboard as you wish.



Q&A