In-class activities: Dec 11, 2017

In our last class for the semester, we will learn about data visualization. You will work on creating useful visualizations of the data you have been exploring through Postgres during the semester. We will be working with Tableau Desktop to create interesting visualizations of the worlddb, imdb, dblp, and lastfm data sets. The software is installed on our classroom computers, but it is also freely available to students:

https://www.tableau.com/academic/students

**Pacing:**

In a flipped classroom, the activities give you some structure, but the objective is that everyone can adjust their pace according to their needs. That means that you are in charge of your pace. This is particularly true for today's activities! Feel free to focus on and explore the visualizations and datasets you find most interesting, and share any interesting findings with your table and the rest of the class.
**Step 1: Tableau Setup**

First, make sure that your postgres server is running, and you can connect to psql as you normally do:

1. Start the psql server
   ```
   sudo -u postgresadmin postgresUmassctl start
   ```
2. Connect to the server
   ```
   psql -U postgresadmin -d postgres
   ```

Make sure that you see the databases we've been working with. We will start with worlddb:

3. Connect to the `worlddb` database
   ```
   \l -> to list available databases
   \c worlddb -> to connect to worlddb
   ```

If everything works OK so far, you are ready to start Tableau:

4. Open Tableau from your laptops Applications folder. (Or find it using spotlight search.)

You need to connect Tableau to the databases we will use (Screenshot 1.1).

5. Under “To a server”, select “PostgreSQL”. If it is not listed, click “More...” and select it there. Fill in the following information (Screenshot 1.2):

   - **Server:** localhost
   - **Port:** 5432
   - **Database:** worlddb
   - **Authentication:** Use only the username `postgresadmin` (you shouldn’t need a password)

*Note:* On your own machines, you need to use appropriate username and password. Also, if port 5432 is not the one psql uses, try `\conninfo` in the psql terminal to check the port number
In case there is no PostgreSQL entry here.
6. You should now see a view similar to the one below (Screenshot 1.3).
Step 2: Tutorial Exercises with worlddb

This step includes three exercises to help you get familiar with the Tableau software. We will do these exercises together. We will be constructing a bar chart, a pie chart, and a geographical map.

Exercise 2.1. Average and total number of people (bar charts)

We will visualize the average and then total number of people in each district of each country in the database.

First, you need to drag the country and city tables into the main view and specify how to join them (Screenshot 2.1). Click the join icon, and select the appropriate fields from the drop-down menus: countryCode for city and code for country.

![Screenshot 2.1](image-url)
Now, click on “Sheet 1”. Drag population (under city) to the columns field and country name and city district to the rows field. To see the average population, select Average in the Measure tab, and to see the total population, select Sum. (Screenshots 2.2 and 2.3). Under Marks, you can choose size, color, and labels of the bars.

Show the total population in each city of the districts. Show Country, District, City and its population. Is it different from the average number? Why?
Exercise 2.2. Number of countries in each of the continents (pie charts)

Suppose we want to see how many countries there are in each continent. In SQL, you would write a query such as this:

```sql
SELECT COUNT(NAME), CONTINENT
FROM COUNTRY
GROUP BY CONTINENT;
```

We will now find the answer through Tableau.

First, change the table you are working on to country (Data Source). Under dimensions in Sheet 1, select the field Continent. From the drop-down, select Create -> Group... (Screenshot 2.4).

You will now see a new field Continent (group) under Dimensions.
Drag the following fields into the designated locations of the Marks box:

- Continent (group) → Color
- Name → Label
- Continent → Label

From the drop-down on Name, set the Measure to count (so you count the countries). See screenshot 2.5.

To select a subset of the continents, drag Continent into the Filters box and make a selection (Screenshot 2.6).
Show how many languages are used in Canada, the United States, China, the Russian Federation, and India. Draw a bar chart showing name of the country and its respective number of languages.

Exercise 2.3. Countries and their cities on the geographical map

Let’s show countries on a geographical map as well as all the listed cities inside of each country. First, join City and Country on countrycode. Then, assign geographical roles to country names and city names (Screenshot 2.7). Now, you can easily show them on the map.

Can you see all countries on the map? Why?
Do you only see US cities on the map? How can you fix this?

Show the population of each city on the map by color density, i.e., the darker the color the greater the population.
Step 3: More exercises with IMDB

This step includes three exercises to help you get more acquainted with Tableau: one for a bar chart, one for a line graph, and one for the “packed bubbles” visualization.

You’ll need to connect to the IMDB dataset for these exercises. You need to connect Tableau to the appropriate data source. Select “New Data Source” from the Data menu. Fill in the fields as in Step 1, but enter imdb2015 as the database.

Note:
You may want to click the “Pause Auto Updates” button on Tableau’s UI when updating your visualization. Otherwise Tableau will try to automatically generate visualizations when you aren’t done selecting the proper fields, and the queries can take a while. Don’t forget to unclick updates once you finish! (If you prefer to have visualizations update in real time, you can just leave Tableau un-paused).

Exercise 3.1: Line graph of movies since 1990

Let’s look into the movie trends of the past few decades. We’re interested in the total number of movies that have been added to the IMDB database for the years 1990 – 2015 (the year the database was downloaded).

You will only need the movie table for this visualization. Construct the visualization with the instructions below:

1. Drag the ‘movie’ table from the section on the left to where it says “Drag tables here”
2. Now change over from the table selection tab (Data Source) to the visualization tab (Sheet 1) on the bottom left of your screen.
3. Set a filter on the number of years to be between 1990 and 2015.
   a. Do this by dragging ‘Year’ into the ‘Filters’ box and setting the range of values to be between 1990 and 2015
4. Drag ‘year’ to the columns field, and drag ‘Id’ into the Rows field.
5. Select the dropdown on the right side of ‘Id’s oval that you just dragged into the rows field, and set the Measure to ‘Count’.

You should end up with a line graph that has two peaks. Are these peaks what you would expect?
**Exercise 3.2: Bar graph of genres**

Now we will find the total number of entries for each genre in the database.

You will only need the genre table to make the right bar chart.
You should only need genre.genre and a count to make the bar graph that you need.

Using your bar graph, what is the most popular genre in the database? Is it far more frequent than the other genres?

**Exercise 3.3: Chris Pratt genre breakdown (Optional)**

Now, let’s take a closer look at Chris Pratt’s acting career. We’re interested in finding what genre(s) of media he likes to act in, and we will be using the “Packed Bubbles” visualization type to do so.

To create this visualization, join the actor, casts, and genre tables.

Notice how you can hover your mouse over the “packed bubbles” button in the “show me” tab on the right and see what is required for this visualization: 1 or more dimensions, and 1 or 2 measurements.

Drag over genre.Genre and count(casts.Mid). Then, once the “nested bubbles” button becomes opaque, click it and let Tableau do the work for you.

Now you should have a nice breakdown of Chris Pratt’s acting career (make sure you have added the proper filters!). Would you say that he’s a funny guy? (Based on these numbers, at least?)
Step 4: Additional Practice

For these questions, we ask that you find the answers with Tableau visualizations only. SQL queries could get you the answers, but that would ruin all the fun now wouldn't it? 😊

- **IMDB**
  - What were Will Smith’s top 3 most common Genres?
  - What were the two most popular genres of 2004?
  - Were *Short* productions more popular in 1913 or 1987?

- **WorldDB (Optional challenging exercises)**
  - Sort the government forms from the most popular to the least popular. Demonstrate with a bar chart and name the top 5 most popular government forms.
  - Calculate gnp per capita for each country and sort them in decreasing order. \( \text{gnp per capita} = \text{gnp} / \text{population} \).
    (Hint: you’ll need to create a new calculated field.)
  - Show the 10 most popular languages in the capitals using a pie chart.
    Note: you will need two filters for this: one to find capitals, and another to choose the top 10 languages spoken in the capital.
Step 5: Group Activity

Now that you’ve had ample practice with Tableau, you can get creative with your visualizations!

Choose any one of the four databases that you’ve worked with all semester: worldDb, imdb, dblp, and lastfm. Connect the dataset to Tableau, and discuss as a team what visualizations you think would be interesting ways to look into the data.

If you have time after you have finished constructing the visualization, ask groups around you what visualization(s) they chose to make and feel free to share your visualizations with each other.