DA 320 Assignment 7 M. Blanco

**Part 1: Short Answers (use 3-5 lines to answer the following questions)**

**Problem 1:** Discuss why and how the ability to create visual representations of data is more accessible than ever.

There are many open source tools like R and Python that allow users to create impressive visual representation of data. There are also many commercial tools like Tableau and PowerBI. These tools provide many graphing features.

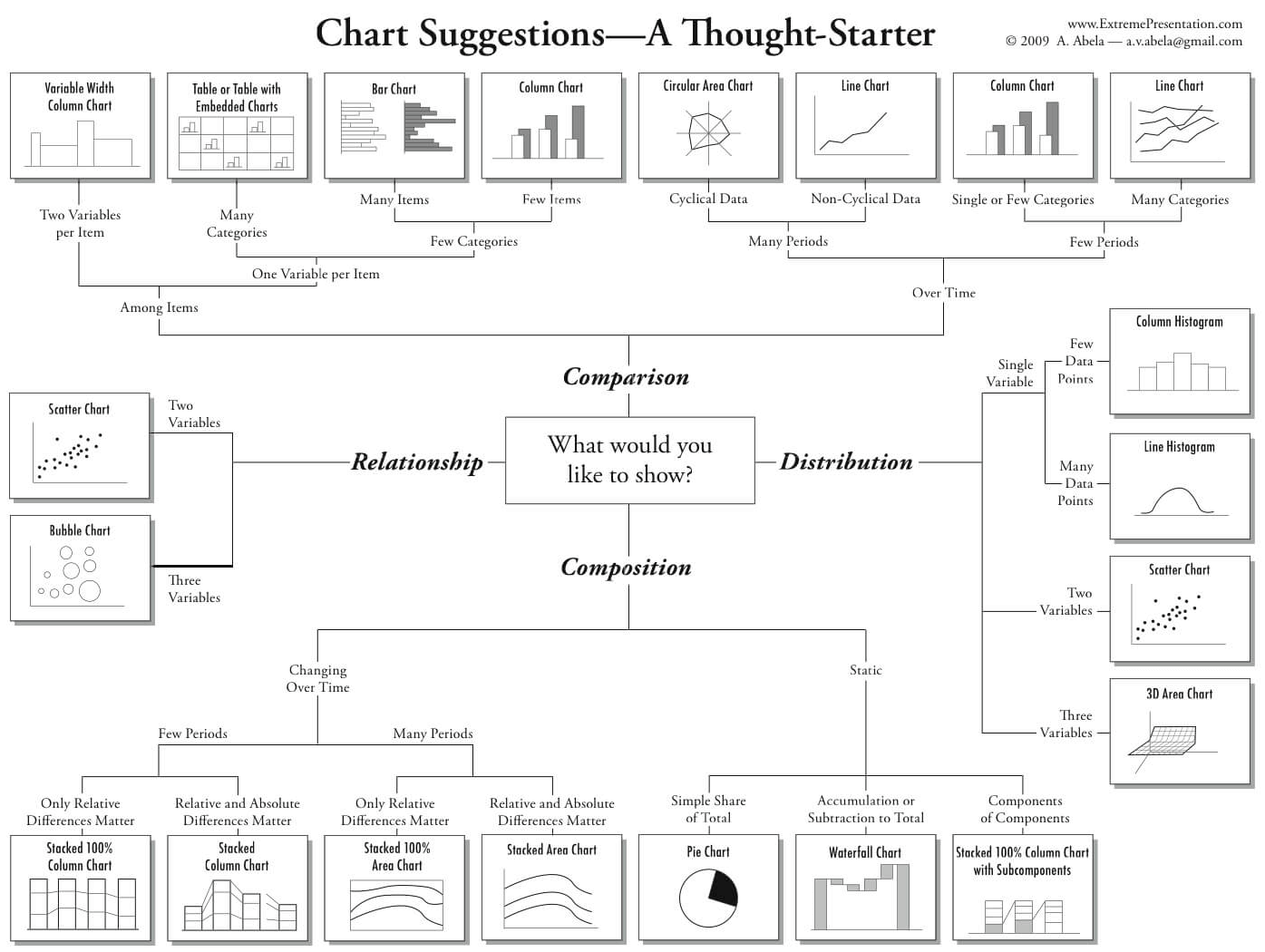
**Problem 2:** Explain how interactive visualizations provide another strategy for providing visual representations of large datasets.

Data visualization has been evolving at a fast pace over the years. Interactive visuals allow users to select planes of data to explore. One recent trend is the growing amount and prevalence of location/geographical data which require an interactive map to fully tell a story with data. Interactive visualization aims to impress with appealing colors and patterns. Interactive visualizations are a blend of data and art.

**Problem 3:** Explain how to decide which visual should be used to best represent the information.

There are four basic visual types that can use to present your data: comparison, composition, distribution and relationship. The following questions can be help determine which visual is best to represent data:

* How many variables will be shown in a single chart
* How many data points will be displayed for each variable
* Will values be displayed over a period or among items/groups



**Part 2: Research Analysis**

**Problem 4:** In regards to human scaling versus machine scaling, humans have scaled to smaller mobile devices such as smartphones and tablets. This creates a challenge for migrating data visualizations to smaller displays. Research and address which visualization techniques help maintain most of their advantages when scaling to both larger and smaller displays.

* Fit
  + Make sure content fits naturally on small screen sizes
* Avoid object overload
  + Limit the number of objects to just those really need
* Arrangement
  + Stack content vertically for small tablet or phone screens.
* Size controls
  + Default, fit all, and fit widt
* Peek
  + Use peek to indicate when more content is available by scrolling
* Geolocation
  + Use geolocation to filter data to what is most relevant
* Target
  + Touch targets need to be bigger than mouse target

<https://www.tableau.com/about/blog/2016/8/tips-designing-device-specific-dashboards-make-everyone-happy-57548>

https://www.tableau.com/about/blog/2016/10/why-mobile-data-visualization-isnt-just-small-viz-61612

https://www.ibm.com/developerworks/library/wa-interface/

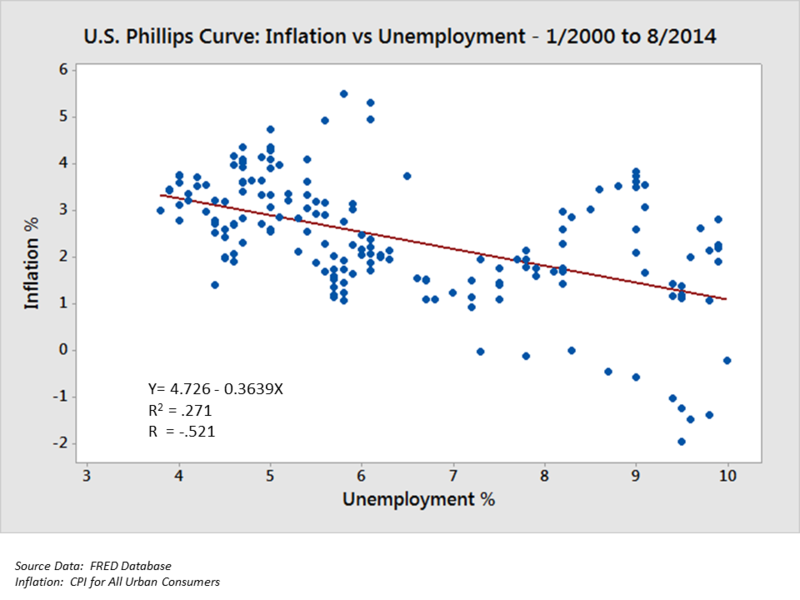
**Problem 5:** Research the three categories of visualization: exploration, confirmation, and presentation. For each category, provide a detailed description, an example with image and URL, and explain how it fits into the category.

**Visual data exploration** aims at integrating the human in the data exploration process, applying its perceptual abilities to the large data sets available in today's computer systems. (Keim) Visual data exploration is used to present the data in some visual form, so user can get insight into the data, draw conclusions, and directly interact with the data.

Starting point: data without hypotheses about the data

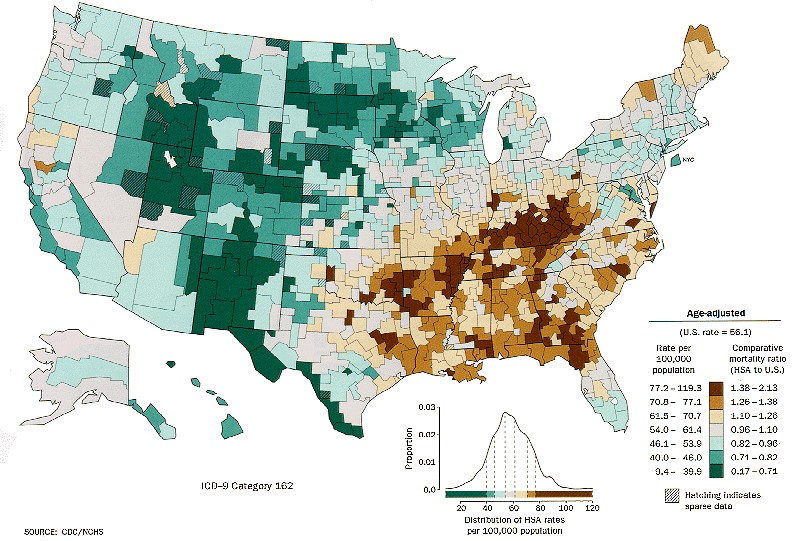
Process: interactive, usually undirected search for structures, trends, etc.

Result: visualization of the data, which provides hypotheses about the data



This is a scatterplot showing negative correlation between inflation and unemployment measured at different points in time. Users can get insight such as negative relationship between the two variables.

URL: <https://en.wikipedia.org/wiki/Exploratory_data_analysis>



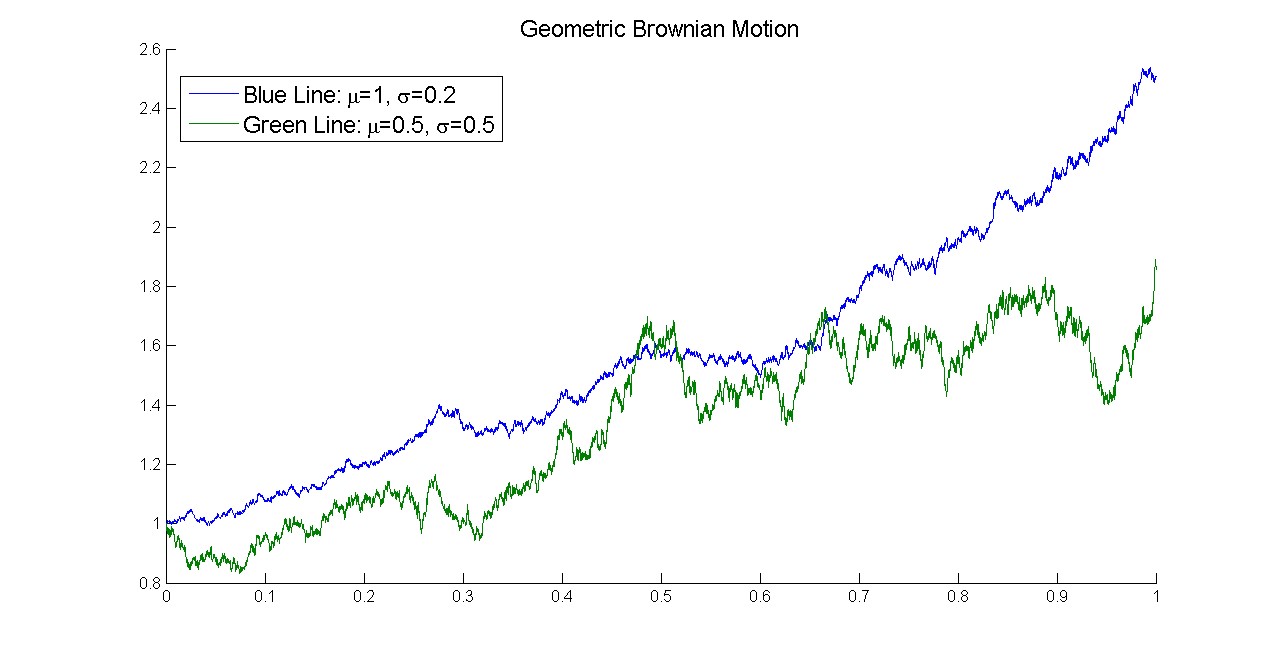
URL: <https://www.interaction-design.org/literature/article/information-visualization-a-brief-introduction>

**Visual data confirmation** enable user to observe whether the underlying system has deviated from the model and assess the risk of the actions about to undertake based upon those assumptions.

Starting point: hypotheses about the data

Process: goal-oriented examination of the hypotheses

Result: visualization of the data, which allows the confirmation or rejection of the hypotheses



This line graph is used to show the similarities in Brownian motion between sets of particles. It can used to question the break in the relationship towards the end of the graph. This graphs help us confirm relation between the geometric Brownian motion.

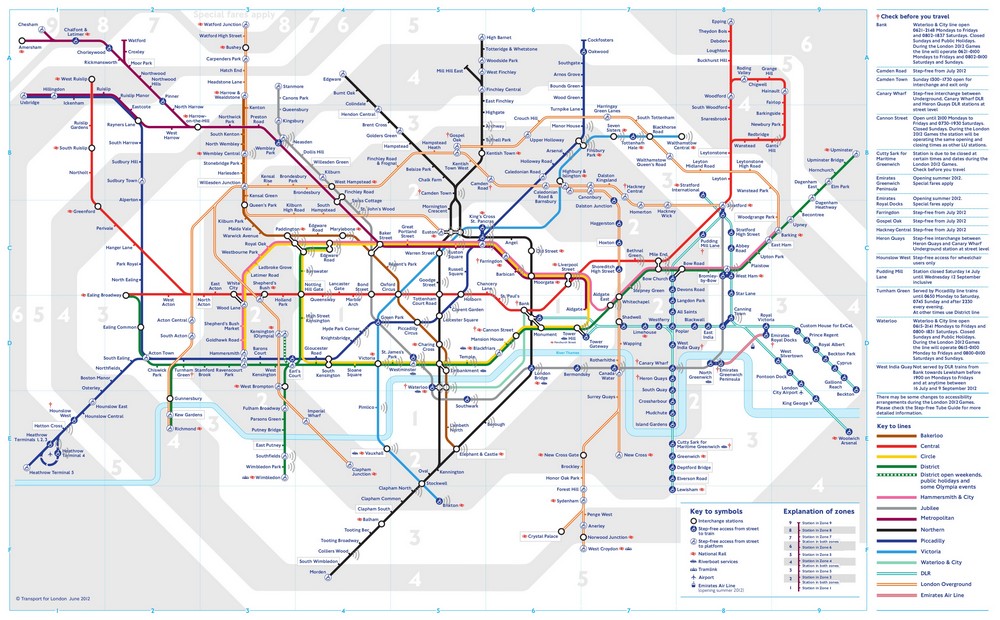
URL: <https://www.interaction-design.org/literature/article/information-visualization-a-brief-introduction>

**Visual data presentation** is used for understanding or persuasion. Visual representation can be used to help understand concepts that might otherwise be impossible to explain.

Starting point: facts to be presented are fixed a priori

Process: choice of an appropriate presentation technique

Result: high-quality visualization of the data presenting the facts



This underground map presents complex data and it’s presented for the purposes of understanding the data.

URL: <https://www.interaction-design.org/literature/article/information-visualization-a-brief-introduction>

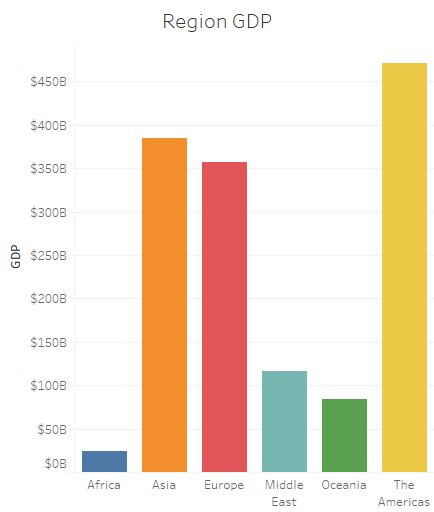
**Part 3: Problem 6:** Follow Tableau tutorial, and create two more different types of graphs, and answer the following questions:

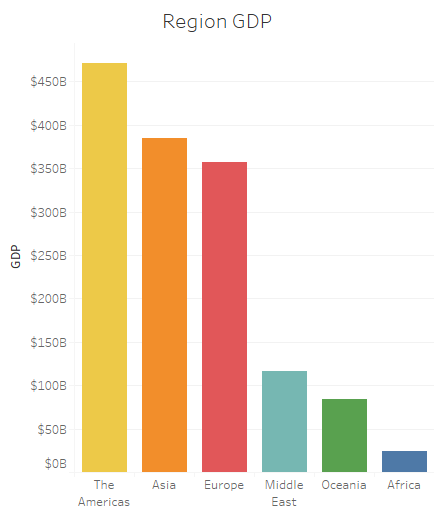
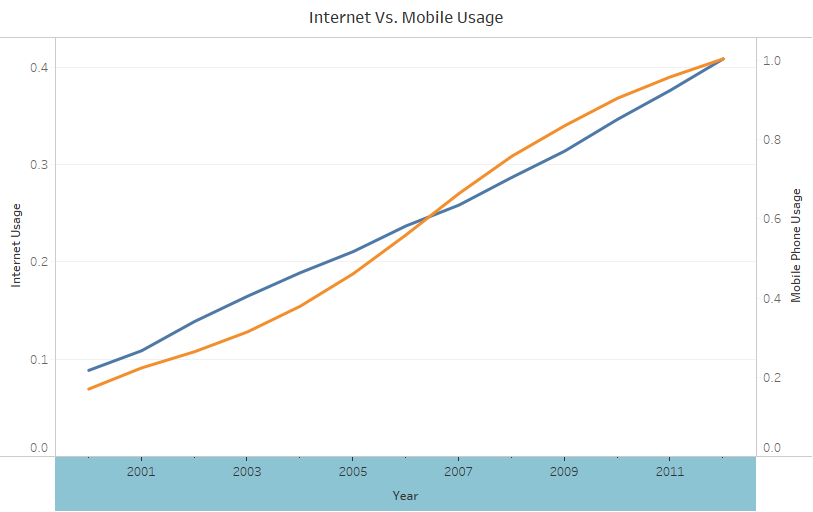
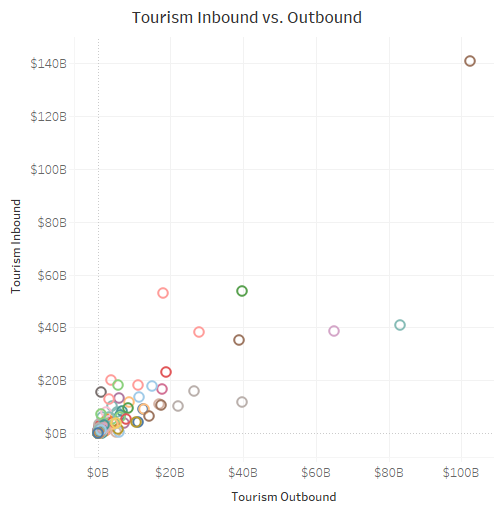
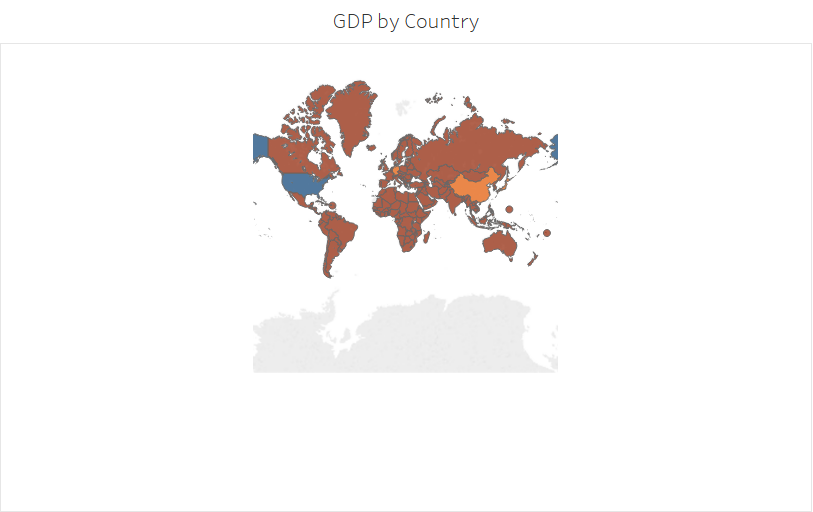
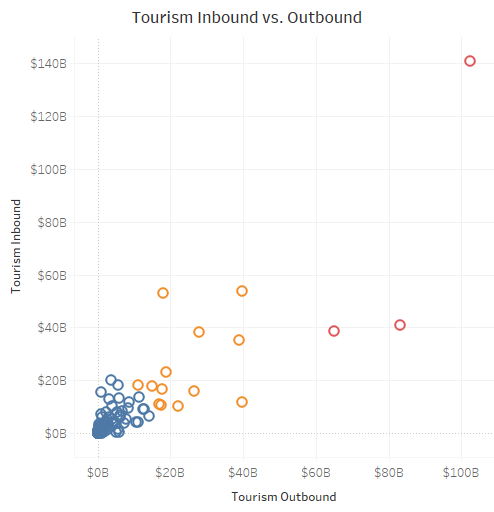
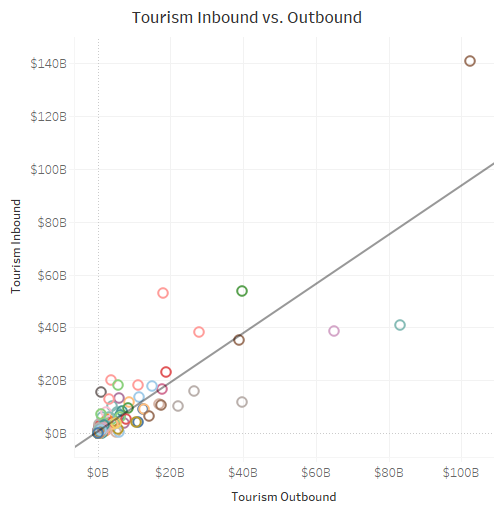
* Why you would like to create this type of graph?

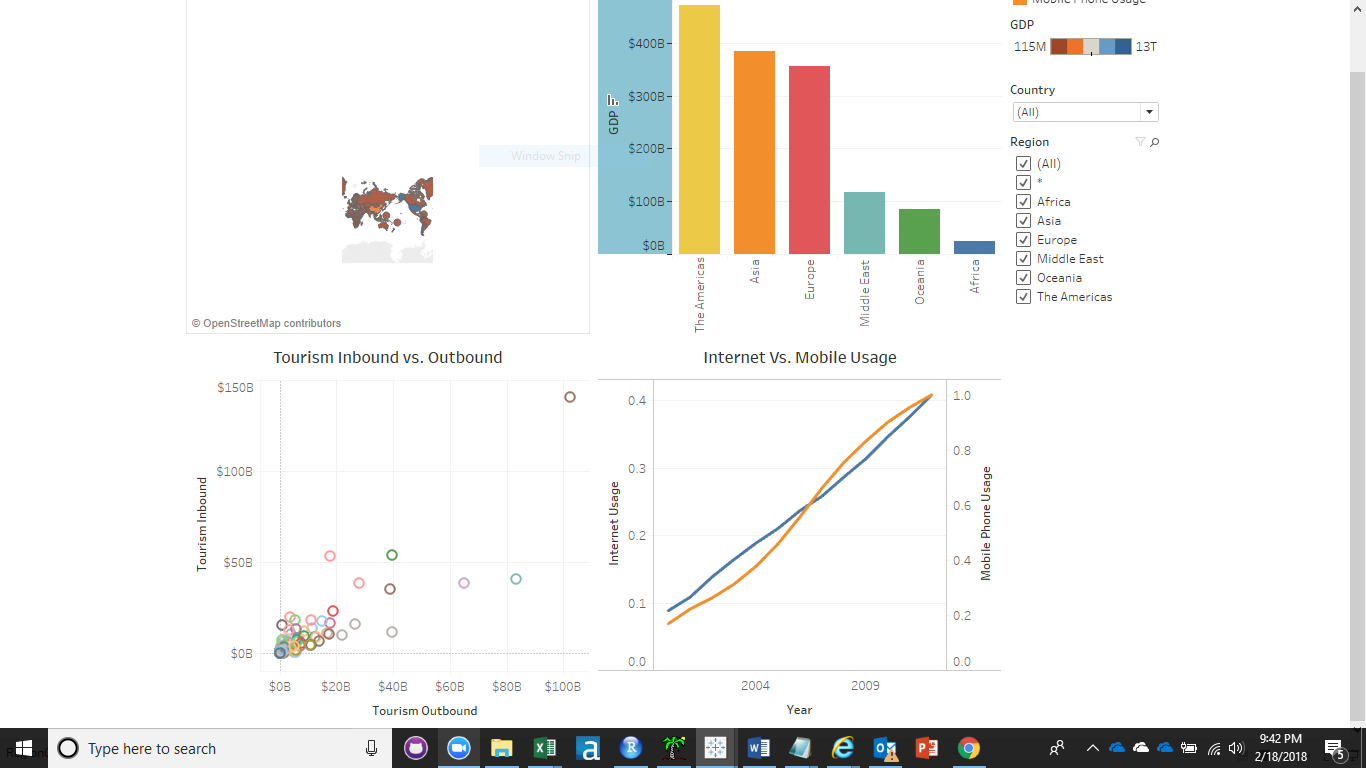
I was interested in creating a line chart, a scatter plot, a bar plot and a map plot.

* What does this graph show? Please interpret the graph as detail as possible.
* Based on your knowledge, is this a reasonable presentation? Why or why not? Explain in details.

|  |  |  |
| --- | --- | --- |
| Graph Title | Description | Reasonable Presentation |
| RegionGDP | This is a bar plot with total GDP for each region for all years. The America’s have the largest total GDP. The current chart is sorted by region alphabetical order but it can also be sorted by increasing/decreasing total GDP. Data can be filtered by year and region. | Yes. Bar plots are good at presenting categorical data with rectangular bars with heights to the values that it represents. |
| Tourism | This is a scatter plot that compares inbound and outbound tourism. There is a positive relationship between inbound and outbound tourism. | Yes. Scatter plot are good at presenting numerical data. Scatter plots are good at showing how much one variable is affected by another. The relationship between two variables is called their correlation. |





**Part 4:** Save your file as ***DA320\_Assignment7\_XXXXX.docx (or .pdf)*** where ***XXXXX*** is the first five letters of your last name, and submit it online.