BIS 449 – Data Visualization
Final Assignment

Introduction

In this assignment, you are required to use both the conceptual and technical knowledge from the course to design a set of informed visualizations. Informed means, you are no longer a layperson in this area! In addition, you must ***explicitly*** discuss the quality of each visualization and your choice and use of pre-attentive attributes; this in addition to very briefly describing what the visual depicts (a good visualization speaks for itself) .

> ***Explicitly*** here meanis you must explain and present the rationale for your design, a rationale that demonstrates underlying concepts from visual design, communication, and story-telling imperatives).

This means this assignment isn't only about producing charts, it is also about reflecting on why they are effective or insightful relative to other or competing choices you could have made or other have made from the same dataset.

These instructions lay out minimum and guiding requirements, you have some flexibility in how you go about completing the overall task, e.g. you could add an additional filter(s)/legends etc. even where it isn't explicitly specified in these instructions. However, you will do so provided your filter/legend etc. adds to insight, chart effectiveness, beauty, truthfulness.

Similarly, these instructions give no explicit prescriptions regarding the use of pre-attentive attributes, you are required to demonstrate your understanding of their effective use by making the right choices and discussing those choices.

The Datasets:
†   **Covid-19 dataset from The New York Times (link)**
> You must download the entire data set, and read the readme (available on the website as well as download). There are three data files categorizing daily confirmed new *cases* and *deaths* for the *US* (national), and by *state* and *counties*. All data files contain a FIPS attribute code, with some geographic exceptions described in the readme file. FIPS codes are a standard geographic identifier like any ANSI standard.

†   **Census Data**
> This dataset has population size (actuals from 2010 census and estimates for future years). The data source is here (bottom of the webpage):

Datasets

Population, Population Change, and Estimated Components of Population
Change: April 1, 2010 to July 1, 2019 (CO-EST2019-alldata)

File Layout [<1.0 MB]

Population, Population Change, and Estimated Components of Population Change: April 1, 2010 to July 1,
2019 (CO-EST2019-alldata) [<1.0 MB]

And for simplicity, I have underlined the dataset to Blackboard to the same folder as these assignment instructions.

Confirmed Cases and Deaths *(New, Total, Total cases per 100 000)*
The Covid-19 data records *new confirmed cases* and *deaths* each day (from Jan 21, 2020). From that, one can infer or compute total number of cases (to date or any date for that matter) as well as total cases per 100 000.

General Notes

†   To complete some of the tasks, you will need to wrangle (clean or modify) your dataset by adding or computing calculated fields. For example, you will be required to compute a moving average, and cases per 100 000 of the population, etc. Depending on your proficiency, it may be easier to modify the Covid-19 datasets in Excel and calculate the derived attributes in Excel the file rather than Tableau. You can also use Tableau of course.

†   You must use design concepts learned during the duration of the course (from Knaflic and Cairo) including what constitutes a quality design and the use of pre-attentive attributes in designing visualizations; how we read information on a page, etc. This means, you cannot simply leave chart design to system defaults (i.e. the machine. This requirement obviously imbues you with some flexibility, use it appropriately.

†   For each chart you design, you must appropriately and explicitly), discuss the design choices you made and why. You must do so in a MS Word/PPT document, by copying each visualization (including the Dashboard), label the figure and discuss it immediately below.

Instructions to Tasks:

In general, your visualizations will be analyzing two key attributes: cases and deaths, separately as well as simultaneously.

1. **Rankings:**
You must create dual axis* charts (bar and circle shapes) that show information both by *state* and by *county*. Your visualizations must appropriately display rankings by state; by counties on the metrics: *new cases*, *total cases*, and *total cases per 100 000* for both *confirmed cases* and *death cases*.

†   Rankings of cases by state; county …
†   Rankings of deaths by state; county …
As shown in class exercise

2. **Maps:**

The data set is from the United States only. Design maps for the data set variously on the key attributes: *confirmed cases* and *deaths*. Your maps must appropriately depict data by *state*; data by *counties* for the metrics: number of *new cases*, *total cases*, and *total cases per 100 000* for both *confirmed cases* and *deaths*

> *Notice*: you have not been given an explicitly instruction as to how many maps to design. You have to make a rational decision that allows you to create the required analysis

3. **Timelines:**

Create timelines or timeseries. For timelines use both *bar charts* and *line charts* You may create multiple exploratory timelines. However, you must show *new cases* (confirmed cases and deaths) and their 7-day moving averages. (The smoothing chart should be a line, and the *actuals* a bar chart). You must do the following:

† Compare the United States overall, with New York State, showing *new cases*/7-day moving average in both cases (over the time period of the entire dataset).
   - Daily new (cases) vs. 7-day Moving Average in the United States (national)
   - Daily new (deaths) vs. 7-day Moving Average in the United States (national)
   - Daily new (cases) vs. 7-day Moving Average in the New York City (or state)
   - Daily new (deaths) vs. 7-day Moving Average in the New York City (or state)

† Use Tableau to <u>forecast</u> and visualize *new cases* and *new deaths* for New York City, or other hotspot (New Jersey, Detroit, Chicago, etc. You can use county-level data here, and consult additional news sources to identify cities or states/counties variously flagged as hotspots). Examine the default model, briefly discuss including, for the city/county you are forecasting identify the forecasted number of *new cases* and *new deaths*.

4. **At your discretion, design another insightful visualization from the data**

Based on your understanding of the data, create an effective and insightful chart that is not specified in these instructions. Make it as beautiful as effectiveness and truthfulness allow.

5. **Dashboard**

Design an exploratory dashboard around the Covid-19 analysis you have just completed. Your dashboard must have a title, descriptive text, appropriate annotations, filters and dynamic sheets. When using the dashboard, the user must be able to display (filter) data by state and region. One should be able to filter between new cases, total cases, and cases/per 100 000; and be able to filter by state, and county.

† Bar graph of new, total confirmed cases; cases per 100 000. (Hint: create a parameter for capacity to display both)
† Bar graph of new, total deaths; cases per 100 000 (apply created parameter)
† Map of new, confirmed cases; cases per 100 000. (use created parameter)
† Map of new confirmed deaths; cases per 100 000
† Ranking of new, confirmed cases; cases per 100 000
† Ranking of new confirmed death; cases per 100 000

You are not restricted to only the above items, there are merely a guide to your dashboard. You must show and configure the necessary (dynamic) filters and legends

Submission Requirements:

You must submit the following

1. A presentation with brief discussion of each visualization, where a visualization is shown, labeled, and appropriately discussed immediately below the visualization. You cannot simply present text in one document and visualizations elsewhere. Failure to do this will earn an automatic zero grade.
2. Notice also how elements of assignment requirements are numbered in this document. You are expected to use these in your documentation!
3. Your Tableau workbook or workbooks
4. Your dataset (this is essential, I don't care that it was assigned to you; you will be making modifications and/or downloading the dataset at different dates times)

NB: If anyone of these above listed items is not submitted, you will earn no points on the entire assignment.

Grading Scale

Item	Points
1. Rankings	50
2.Maps	40
3. Timelines	80
4. At your discretion	40
5.Dashboard	90
Total	**300**

NB: Failure to provide a discussion narrative for each visualization with earn no points for the related item regardless of visualization. Your discussion must discuss the quality of the visualization and your choice use of pre-attentive attributes, in addition to briefly describing what the visual depicts.