“Truth and Lies” in Visualization Infographic

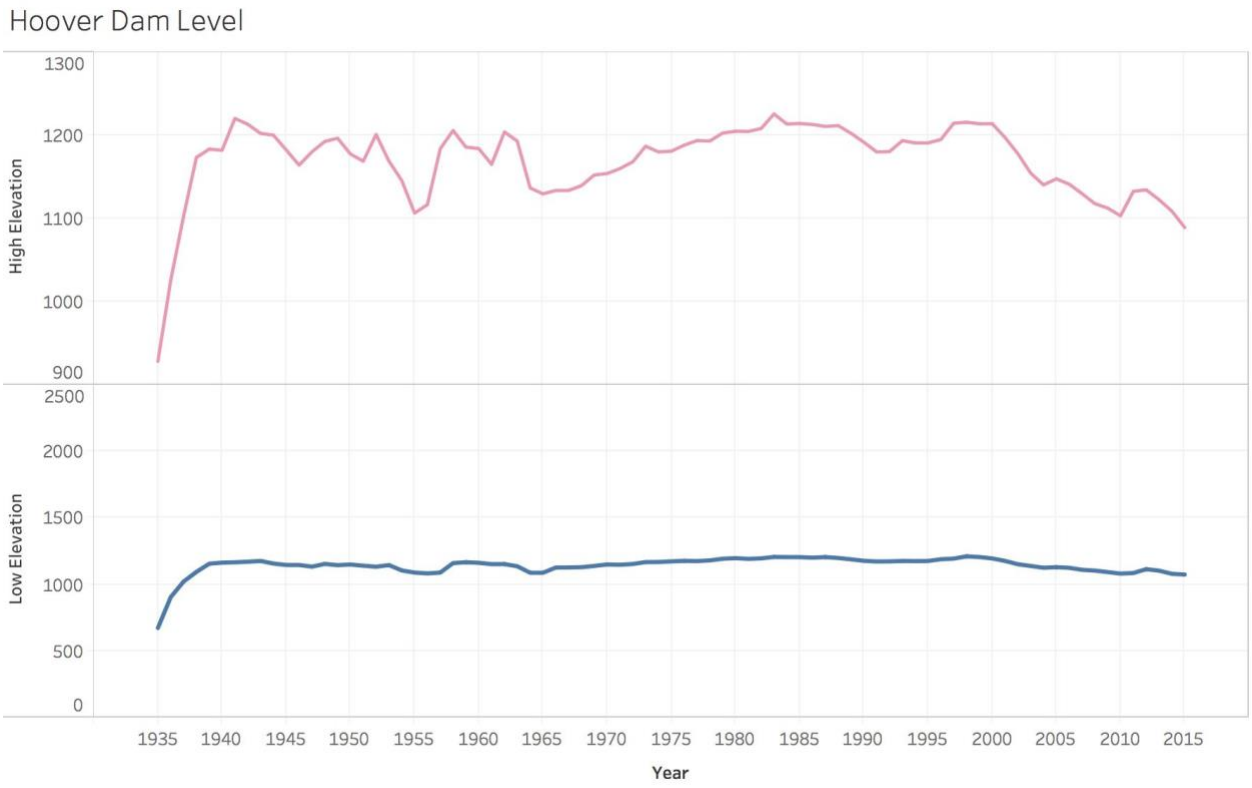
Dataset description

Hoover Dam Data.xlsx is provided by instructor. The title is “Lake Mead and Colorado River Watermaster Hydrographic Daily Log”. The file consists 7 columns and 85 rows, including Year, Date, Time, Low Elevation, Date, Time, High Elevation. From data explore analysis, it is obvious that there are two measures of water level, starting from 1935 and ending in 2015.

Infographic and Data Analysis

In this part, I am going to show the visualization “lies” could lead to misunderstanding and may manipulation audience to have wrong conclusions. I will start with making lies and then showing the truth.

1. Lies



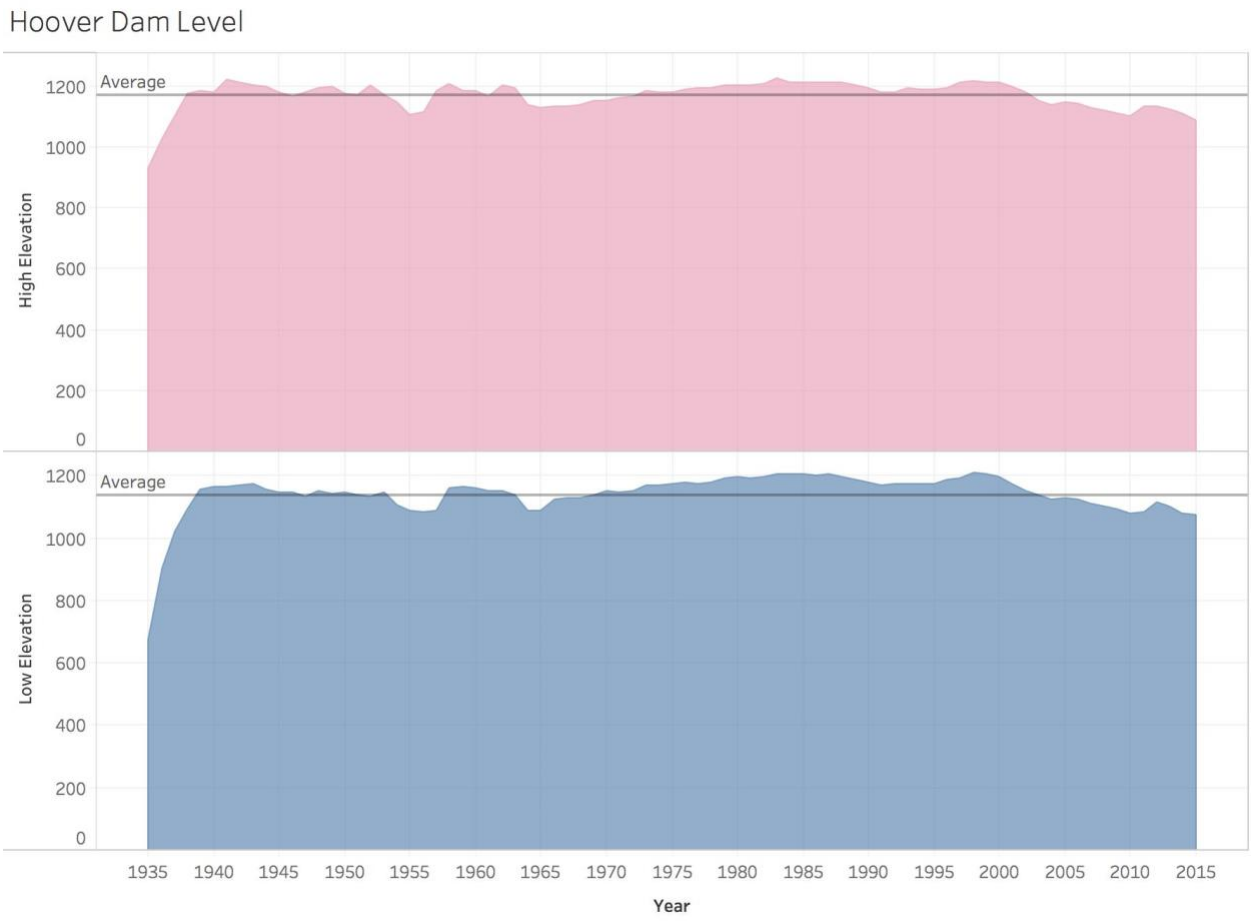
The trends of High Elevation and Low Elevation for Year.

From this image, which shows the water level of Hoover Dam. Pink line stands for High Elevation and blue line stands for Low Elevation. Time range is Year 1935 to 2015.

If we just take a quick look of this infographic, many people may quickly reach to a conclusion that the fluctuation and volatility level of High Elevation during this time period is more drastic than Low Elevation. Since the pink line has more up-and-downs, the blue line tends to be moderate and flat.

However, this is a lie built by this infographic. Though they have same interval in the x-axis (Year), High Elevation and Low Elevation use different range in the y-axis. High Elevation is from 900 to 1300 with 100 unit as a gap, while Low Elevation is from 0 to 2500 with 500 unit as a gap. In this way, the change of High Elevation will be more dramatic compared to the Low Elevation. This is a trick I played by stretching and shrinking scales.

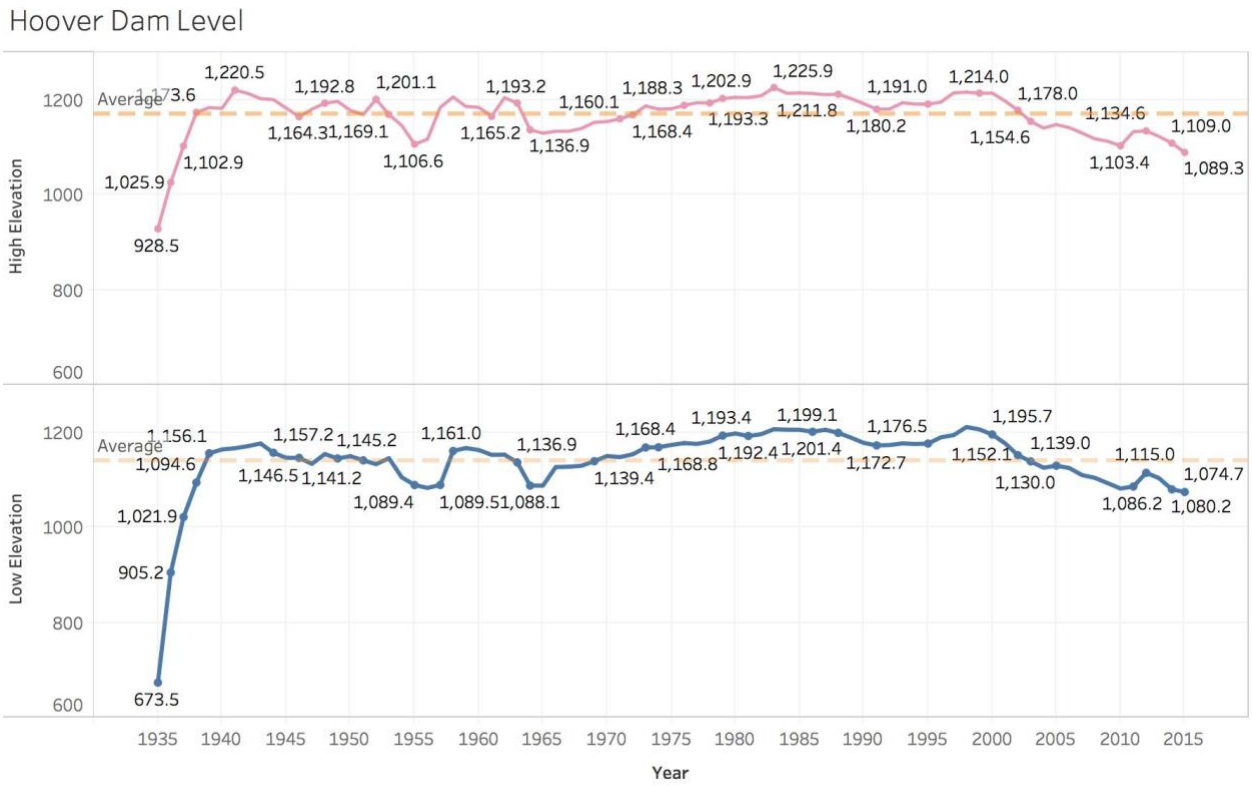
2.Truth



The plots of High Elevation and Low Elevation for Year.

In the area graph, with same range and unit interval gap for Low and High Elevation in the y-axis, we can see that the fluctuation High Elevation is not drastic or violent compared to the infographic in the “lies” part.

Below is another truth visualization, which I enabled to show mark labels for every water level number. Numbers, as a quantitative method in data visualization, will give us the straightforward information. I also added an orange dash line for average water level for both High Elevation and Low Elevation. We can see from both truthful graphs that the water level of High Elevation is not so unstable, but how fluctuation it is compared to Low Elevation still need to be explored.

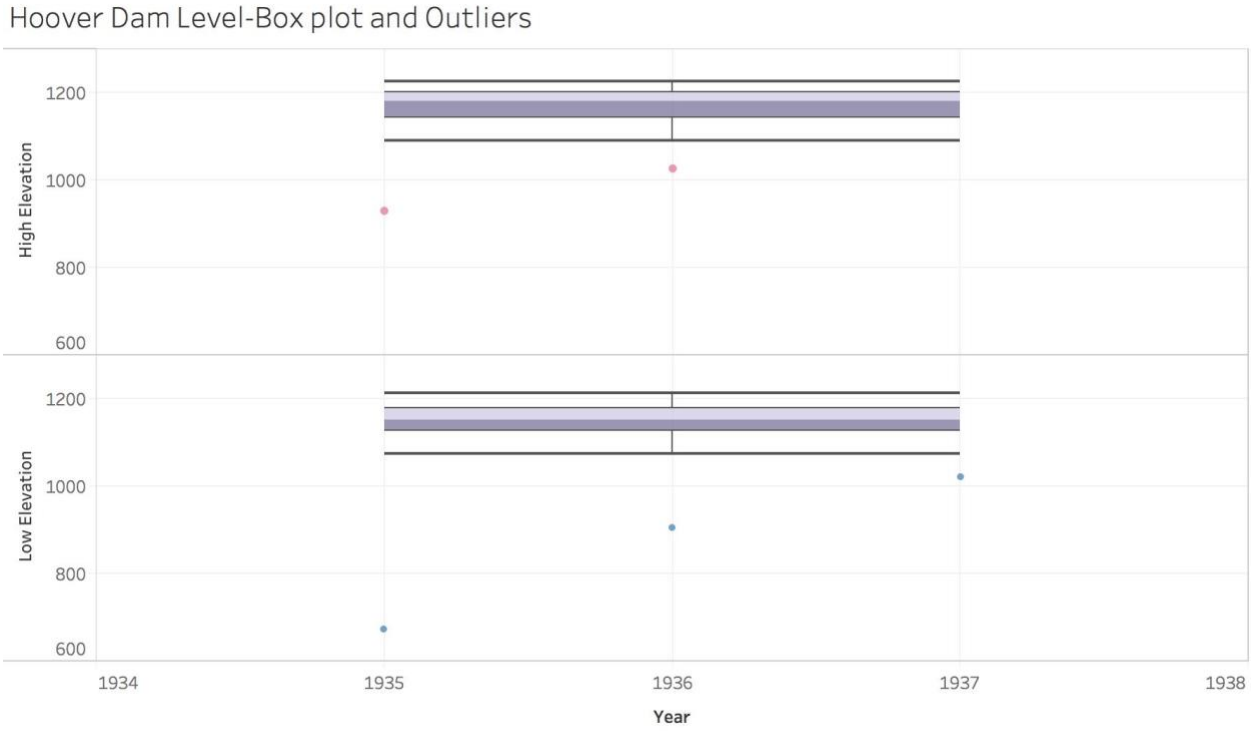


The trends of High Elevation and Low Elevation for Year.

In order to further proof the truth, and find out which Elevation has higher fluctuation and volatility rate, I calculated different levels of measurements in Excel and plotted in Tableau.

	Low Elevation	High Elevation
Maximum	**1211.37**	**1225.85**
Minimum	**673.5**	**928.45**
Annual Average	**1141.052593**	**1169.127901**
Standard Deviation	**71.19125309**	**46.40836309**

From the chart table above, we can see that as a matter of fact, Low Elevation is more unstable compared to High Elevation, since it has higher standard deviation.



The plots of High Elevation and Low Elevation for Year.

Obviously, there are many outliers would be the reason of that. Therefore, I drew the box plot and keep the outliers in the graph. We can easily notice that Low Elevation has more outliers than High Elevation and the distances are far from median, average and its box plot.

In a nutshell, Low Elevation is more unstable than High Elevation for Hoover Dam water level from 1935 to 2015.