COMP30750Visual Exploration Tool Design Document

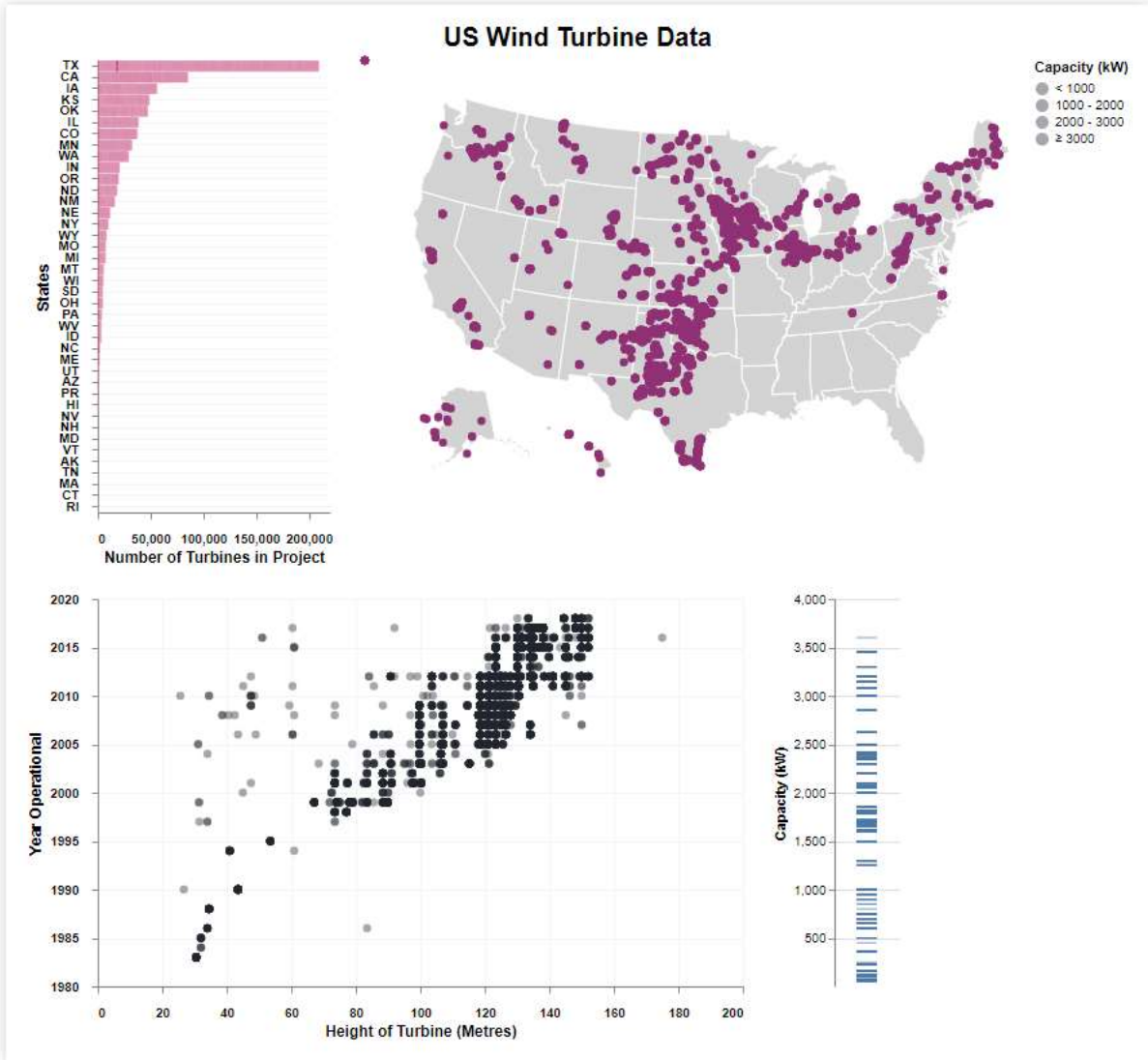
Student Names: Ruth Adesanya, Toluwabori Akinola

Student Numbers: 20322406, 20720321

Title:

Us Wind Turbine Data

Screenshot:



Dataset overview:

This is an adaptation of a dataset from rfordatascience on GitHub. The original dataset provides details on all wind turbines across the USA.

Features on wind turbines available include:

- location (latitude, longitude)
- state of wind-turbine
- county of wind-turbine
- project name
- the year the project became operational
- the number of turbines in project
- the turbine original equipment manufacturer
- the turbine models
- the turbine capacity (kW)
- the turbine total height (meters)

The original dataset provides a vast amount of data concerning all the wind turbines in the USA.

We used Python to derive a new version of this dataset with an ordinal variable that specifies the ranges of our 't_cap' feature. Our new column adds rows based on their wind turbine capacity value and which range they are in. These new column allows us to view data according to which range of wind turbine capacity they are in. We removed a lot of data, that we believed to not be relevant for this analysis. We also removed any wind-turbine data that had missing data as they would lessen the accuracy of the analysis. We also reduced the size of the data we were analysing as Vega-Lite does not work well with a very large amount of data. The shortened csv file we uses is called "wind_usa_cleaned_st2.csv" and is included in our submission.

Design considerations

This should provide an overview of your visualisation, a discussion of why you used specific encoding / interaction options, and the pros/cons of your visualisation vs alternatives.

Overall goal: Our overall goal with this tool was to enable the exploration of the relationship between the height of turbines ,the year the turbine became operational, the state the wind turbine is in and the number of turbines in the project.

Bar chart: This is a bar chart that displays the total number of wind turbines in each state. The bar chart is sorted based the total number of wind turbines in each state. Sorting by this attribute makes it easy to see which states have the most wind turbines but makes it harder to look up a specific state. Although by being able to click and drag, it is slightly easier to see the data of an individual state. An alternative would be to scale the bars logarithmically – this would make it easier to find specific counties, but harder to compare the states in the list. Instead of a normal bar chart We could have used grouped bar charts or small multiples to visualise the different county types. However, We felt that it was sufficient to be able to compare total number of wind turbines per state without needing to accurately compare specific counties.

Tick Graph: This is a tick graph that shows the varying capacity of wind turbines in the USA. We initially attempted to use bar chart, but this suffered from having too many different varying data. The advantage of this tick graph is that it is easier to select certain ranges of wind turbine capacities. The disadvantage of the bar chart was that it was difficult to select ranges in the capacity. We also found that it was not necessary to plot the capacity of the wind turbines against a nominal data attribute.

Dot plot map: This is a dot plot map that shows the location of the different wind turbines throughout the USA. Each wind turbine is represented by a dot that is scaled according to the capacity of the turbine. This approach makes it easy to see the geographic distribution of wind turbines across America and also makes it immediately obvious that there are more wind turbines in the eastern parts of America. A disadvantage of the dot plot map is that it suffers from overplotting, especially in Texas and California where there is a lot of overlapping points. One solution would be to aggregate the wind turbines into a state level summary and show this as a single sized circle instead. The aggregation would prevent inspection of individual wind turbines, which We have enabled here using tooltips. We used a state TopoJSON map of the USA to create the map.

Scatter plot: The scatterplot shows the correlation between the Year the project was operational and the Height of the turbine that came from such project. The intention behind this plot was to see if as the years went by, the height of the turbines increased. Which it did! The graph shows a positive correlation between the year of operation and the height.

This result makes sense seeing as though the need for wind turbines to generate a lot more power increased. And we know that generally, taller wind turbines capture a lot more energy . This is due to the fact that wind increases as altitude increases**.**

From the plot we can also see that majority of the turbines are centred around the 120-150m height mark. This range seems to be the most common heights of a wind turbine. If we hover over the points also, we see that manufacturers tend to make wind turbines of the same length. Regardless of the year it's being operational

Interaction consideration: The main interaction approach We have used here is cross-filtering. Users can select subsets of the data in one chart and this will filter the data in another chart, e.g. users can select only one state in the bar-chart and this will filter the data presented in both the strip plot,

scatter plot and the dot-plot. Each chart allows selection on a different data attribute – the bar chart allows selection by state, the strip-plot allows selection based on wind turbine capacity, the dot plot map has an interactive legend to allow selection based on turbine capacity range, the scatter plot has a click and drag feature that lets you select multiple wind-turbines in the scatter plot. When combined these allow the user to explore the relationship between these nine attributes.