

**Edward Tufte - Visual and Statistical Thinking**

Exercise 1

**Introduction**

“Making decisions based on evidence requires the appropriate display of that evidence. Good displays of data help to reveal knowledge relevant to understanding mechanism, process and dynamics, cause and effect.” E Tufte

The technique used to analyse and display data when assessing a problem can have a significant impact on the results found, as each method is more appropriated for a particular type of data to be analysed than others.

In the paper “Visual and Statistical Thinking: Displays of Evidence for Making Decisions”, Edward Tufte discusses about two stories to show how the choice of methods to analyse and display of evidences could help to make good or bad decisions to solve a problem and understand how things work.

**Cholera Epidemic in London in 1854**

A serious epidemic problem in August 1854 in London would become famous not for the medical approach, but for the data analytics methodology applied by the designated doctor, Dr John Snow, who was eager to answer questions as where the source of contamination was and also find arguments to prove that the water was the vehicle to spread the disease and not air as people believed during that time.

His suspicions were over the water from a communal well in central London, at Broad and Cambridge streets, which he believes was contaminated and motivate his work of collecting evidence on cases around that neighbourhood.

Initially, it was used a time-series plot to display the data, the number of deaths for each day, which shows the chronology of the epidemic, when it started, and its development over time, and helps to understand if the epidemic is getting worse or improving. However, this type of visualisation does not help to understand the origin and cause of the epidemic.

Then, John Snow collected a list of 83 deaths from cholera from the General Register Officer, and plotted that information on a map along with the location of the 13 community pump-wells on that area, which showed a significant correlation between one of the pump-wells (Broad Street pump) and the location where most of the deaths took place. There were a few cases where the death occurred in other area, but John Snow could figure out through some interviews that those people were on the contaminated area and had drunk the water from Broad Street pump.

After a long period of collecting and analysing all these evidence, John Snow presented the result of his findings to the local authorities that decide to take his advice and removed the pump-handle. And then, again the time-series plot could be used to verify if that action had some impact on the number of deaths by cholera.

**Decision to Launch the Space Shuttle Challenger in 1986**

Evaluating Tufte’s report it can give a first misunderstood impression that the author is arguing that the engineers from Morton Thiokol team were responsible for the tragical launch of the space shuttle Challenger in 1986. Although, following an in-depth analysis of his arguments it is possible to evaluate that there were two main points which collaborated to the refute of the engineer team call for delay the launch by NASA:

1. The unclear display chose to demonstrate the data and,
2. The fault in demonstrate a cause-effect relation between the parts analysed with the local temperature.

On the morning of the planned launch day, the weather was colder than foreseen which made the engineering team decided to recommend stopping the flight. The engineering team presented a complex group of 13 different charts, considering that the team was right about its conclusions to support the decision to delay the launch, the inability to express such information in a clear and related way to the authorities audience was directly responsible for make this work to be discredited by NASA, and therefore, not to follow the no-launch recommendation.

When re-visiting the charts, Tufte explains that one of them was showing the immediate danger to the space shuttle through previous records of O-rings erosion in earlier launches, on a scale from simple to catastrophic. As strong as this argumentation could be to highlight how risky it could be for the flight, it still does not relate it to the possible cause, the temperature. Despite not being pictured in the presented set of charts, the temperature was part of the team analyst’s data and more than that, it was a central point of attention in their arguments. However, they had only one similar case to show as a comparison of evidence, which was not enough to make their statement reliable about the cause and effect. As Tufte explains, the showed data makes it clear that the engineering team had all necessary data but was focusing in show it all in individual plots, most of the time too technical and relegating the probably high point, a relation of cause-effect between their findings about the O-ring damage historic and the impact of cold temperatures over it.

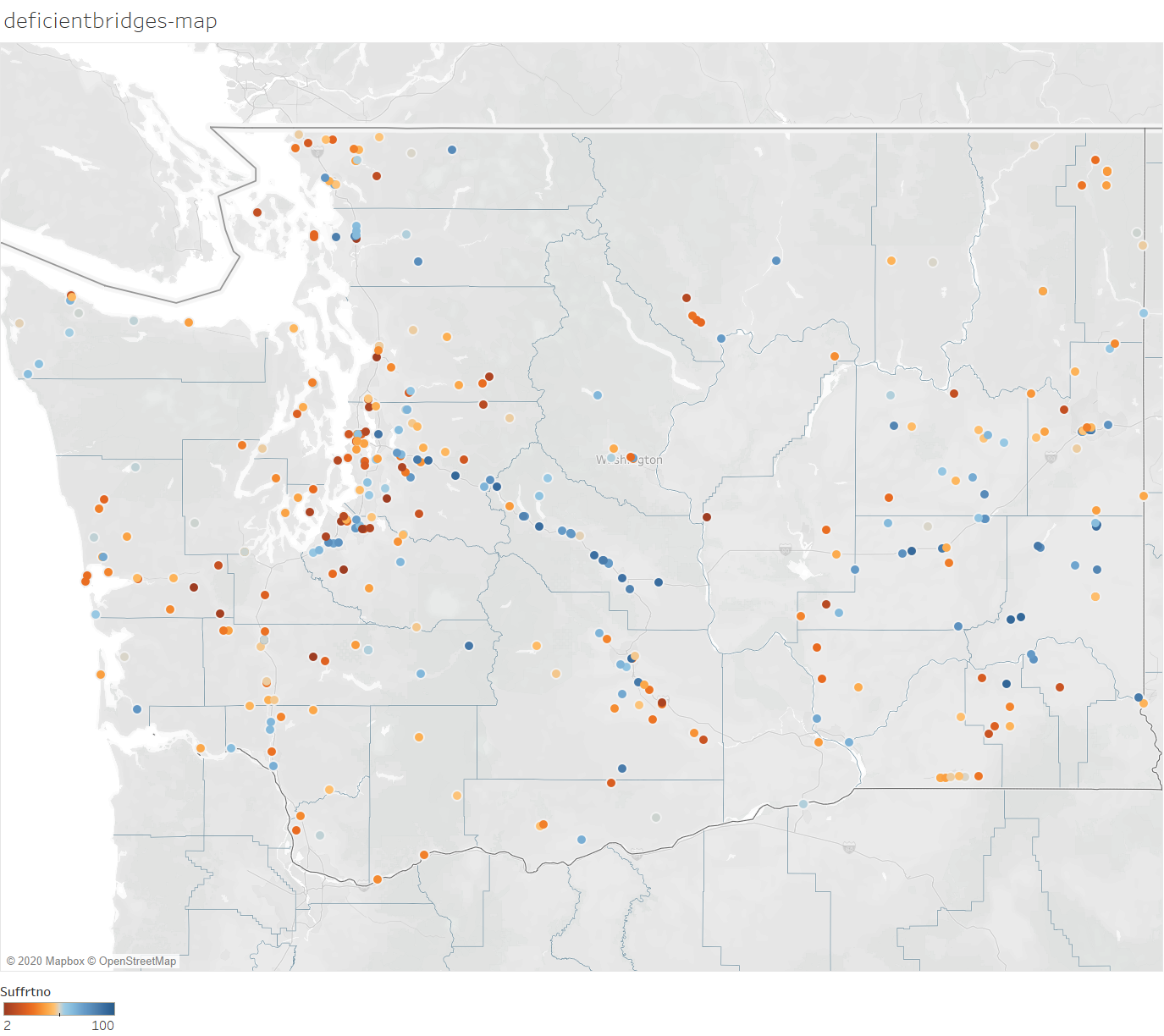
**Conclusion**

For the studied cases in this work, a common conclusion would be aligned with a famous Leonardo da Vinci quote: “Simplicity is the ultimate sophistication”. The capacity to communicate information in reference to data, mainly about statistical data, is directed linked to how clear is the chosen method to display that data in a way the target audience can truly understand and find the right conclusions.

**Tableau Academia: Tableau Data Journalism.**

Exercise 2

When verifying only Structurally Deficient bridges, none of the Skagit River bridges are shown on the map. However, it is possible to notice a significant number of other bridges that need attention as they have the sufficiency rating below 50 (eligible for replacement).

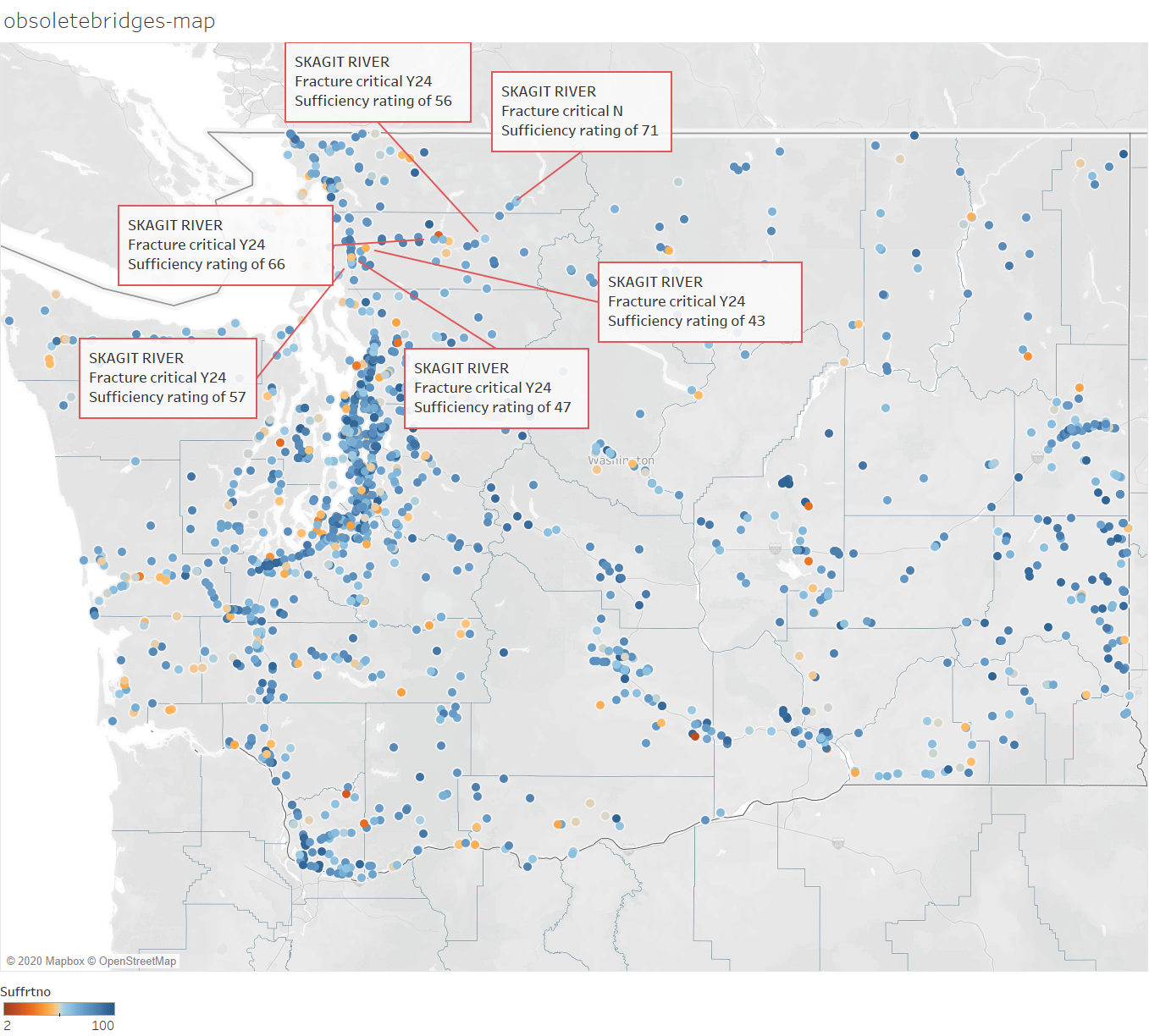


The next board is a heat map by county showing where the largest number of bridges are that require attention due to the low sufficiency rates, and Skagit County is not one of the worst locations.

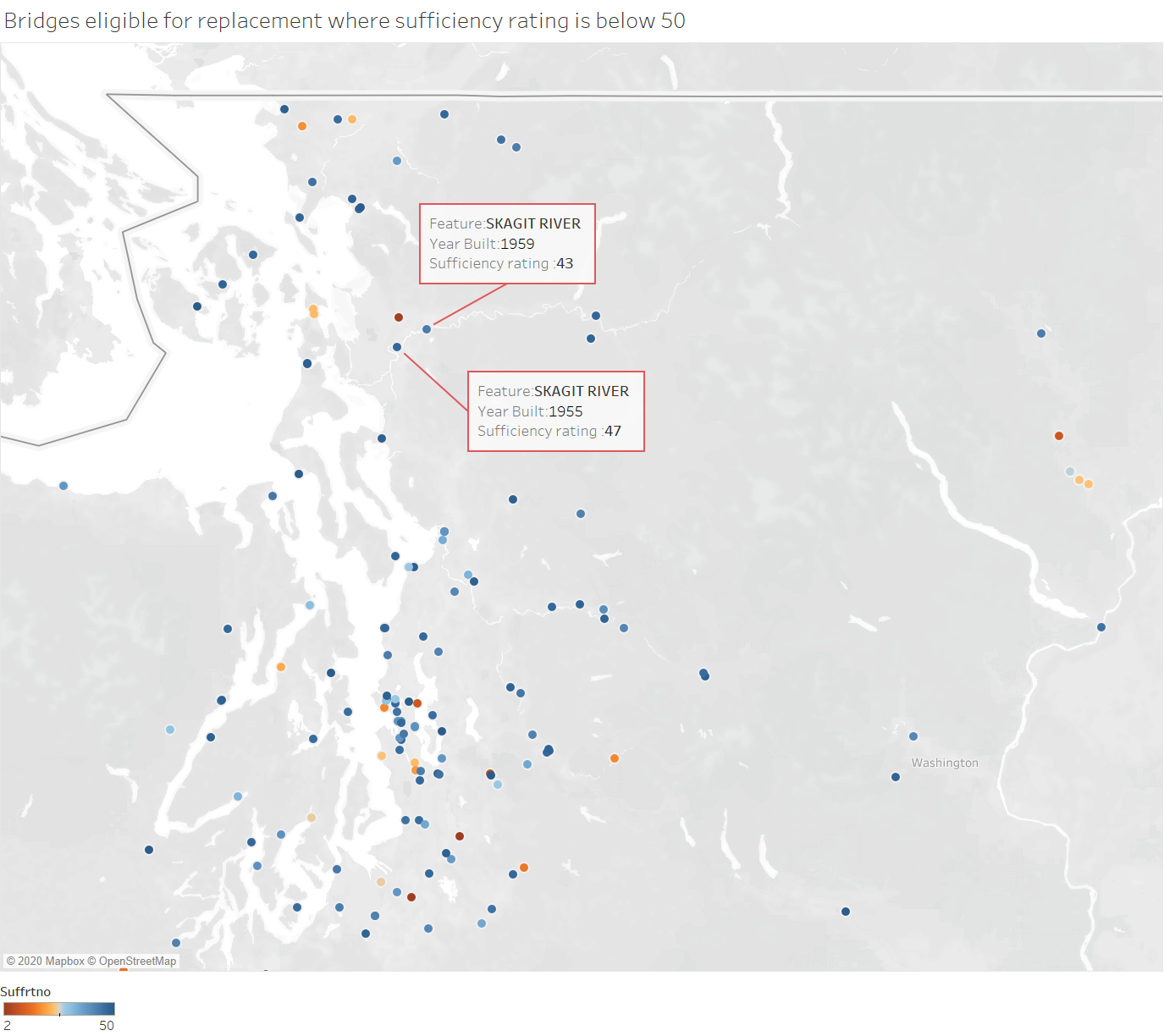
Map

Description automatically generated

Next, I changed the view to show only Functionally Obsolete bridges and then I can see the Skagit River bridges, and I can also notice that most of them are described as Fracture Critical and have the sufficiency rating below 50.



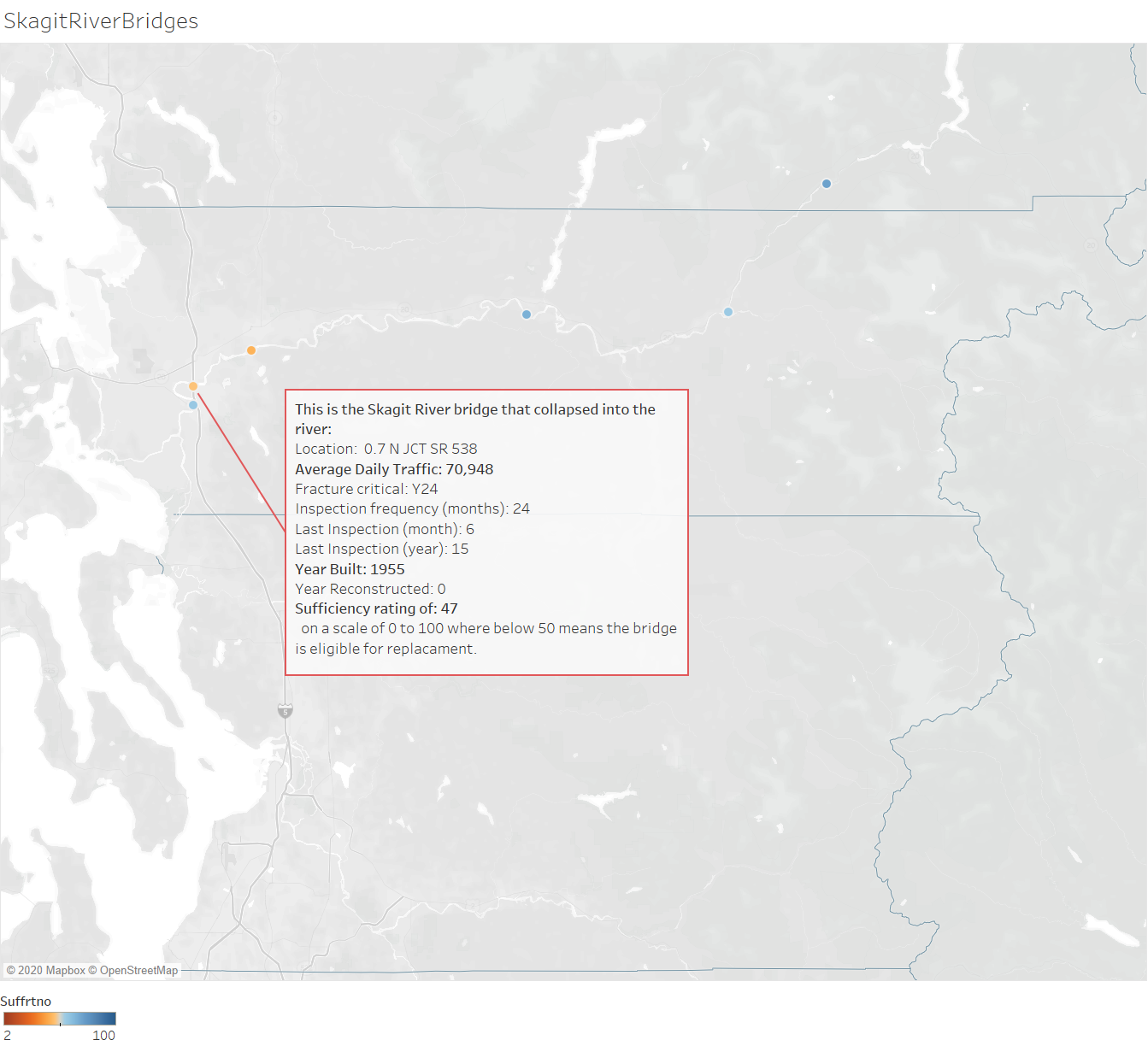
Then I selected only bridges eligible for replacement, where sufficiency rating is below 50, and where the bridge was still open to traffic, and I can now see only 2 of the 6 Skagit River bridges on the map.



Then, I added a filter by Feature to show only the Skagit bridges and based on some extra information from the news posted on The Seattle Times [1], I could figure out that the Skagit River bridge that collapsed was the one pointed out in the next board, as it is the only built on 1955.

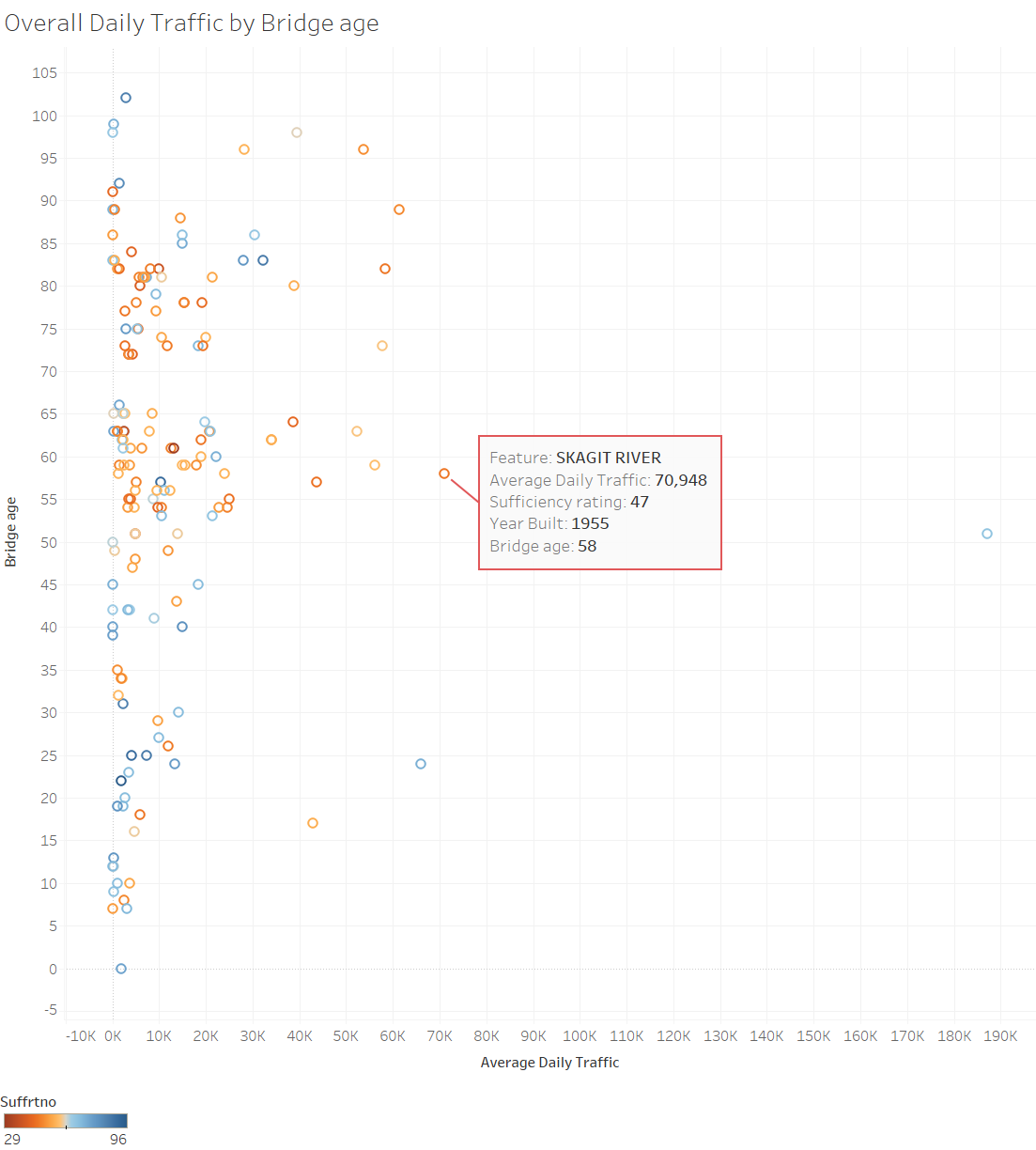
It has sufficiency rating of 47, however, it is not the worst rating for the bridges on that river, and it is described as Fracture Critical as it is an old bridge (1955).

The information available on the dataset shows the last inspection was on 06/2015 but the bridge collapsed in 2013, thus, there is no information when it was the last inspection before the collapse. The bridge was inspected each 24 months.



The average daily traffic was high on that bridge (70,948) when compared to other bridges where the sufficiency rating is higher (see Overall Daily Traffic by sufficiency rating board below). The below board also shows that there are a couple of other bridges on similar situation (sufficiency rating is below 50, a high average daily traffic, same generation or older and the structure were not reconstructed) that might require attention.





References:

[1] Span wasn’t built to take critical hit (2013) The Seattle Times. Available at: <https://www.seattletimes.com/seattle-news/span-wasnrsquot-built-to-take-critical-hit/>

(Accessed: 25 Oct 2020)

Exercise 3

Covid-19 Dashboard

