

**Data Visualisation – CA1**

**B9DA104**

**Master of Science (MSc) in Data Analytics**

**Tim Browning**

**10543928**

**Dublin Business School**

Contents

[Q1) Identify and critique the visualisation techniques that were applied in the two cases to communicate the information necessary to assist decision-making. 3](#_Toc55143495)

[The celebrated investigation of a cholera epidemic 3](#_Toc55143496)

[The unfortunate decision to launch the space shuttle Challenger 6](#_Toc55143497)

[Q2) Build a story board to show deficient bridges (slide 17) and explore the unsafe bridges (slide 20). Build the story around the causes of the collapse of the Skagit River Bridge. 10](#_Toc55143498)

[a) What’s the rating of the Skagit River bridge that collapsed into the river? 10](#_Toc55143499)

[b) What other distinguishing information can we obtain that might go into a story? 10](#_Toc55143500)

[c) When was the bridge last inspected and how often is it inspected? 11](#_Toc55143501)

[d) Is it considered fracture-critical, meaning one hit to a span can knock it out? 12](#_Toc55143502)

[e) How much average daily traffic does it get? 12](#_Toc55143503)

[Q3) Build a Tableau COVID-19 Dashboard including three visualizations telling a story related to the pandemic? 13](#_Toc55143504)

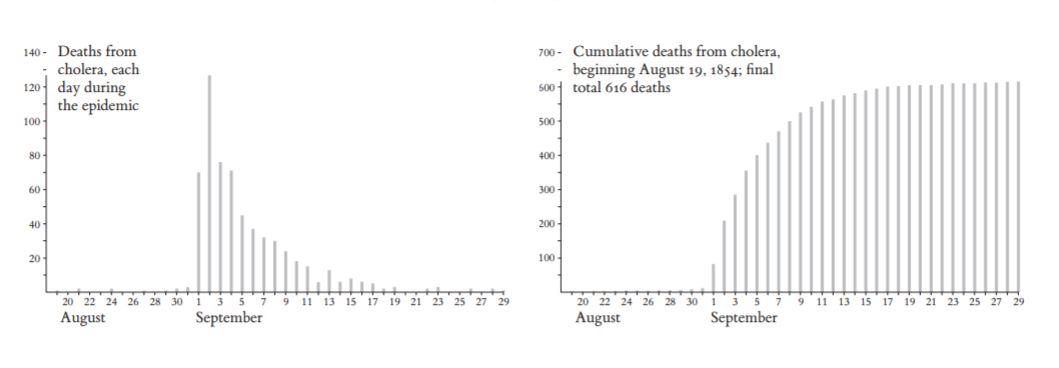
# Q1) Identify and critique the visualisation techniques that were applied in the two cases to communicate the information necessary to assist decision-making.

Data visualisation is the graphical representation of information and data.

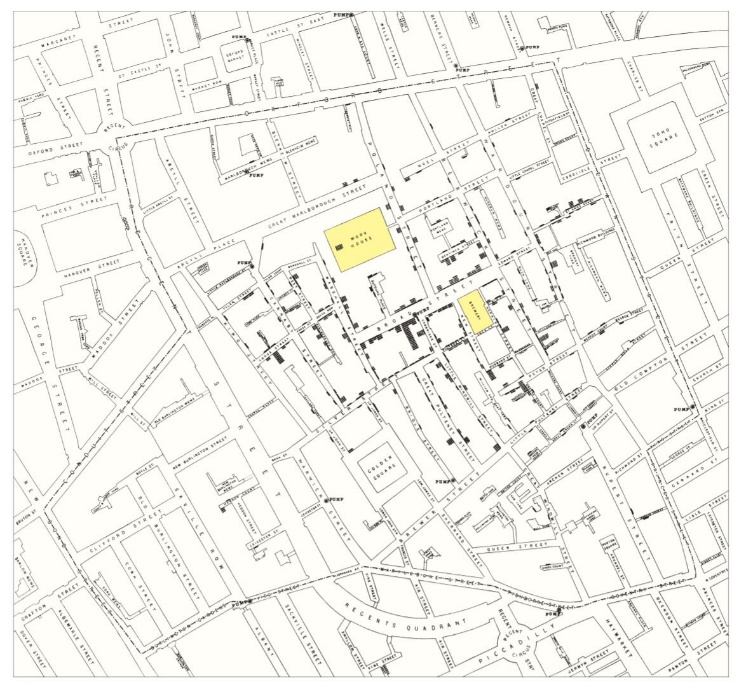
The two cases presented in this study are:

The celebrated investigation of a cholera epidemic and the unfortunate decision to launch the space shuttle Challenger

## The celebrated investigation of a cholera epidemic

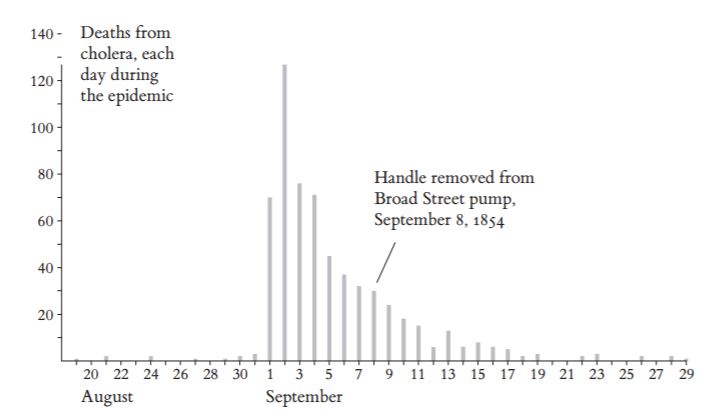
For assessing cause and effect they first created a **time-series graph**. The passage of time was deemed to be a poor explanatory variable and useless in finding how to intervene and stop the pandemic. This graph would report each day’s deaths and would give no insight into the location of where the epidemic is most prevalent.

Snow created a **map** which marked the deaths from Cholera as well as marking the location of each water pump. One well in this photo below is located where there is a cluster of deaths. This map reveals a strong association between Cholera and the location of the well.



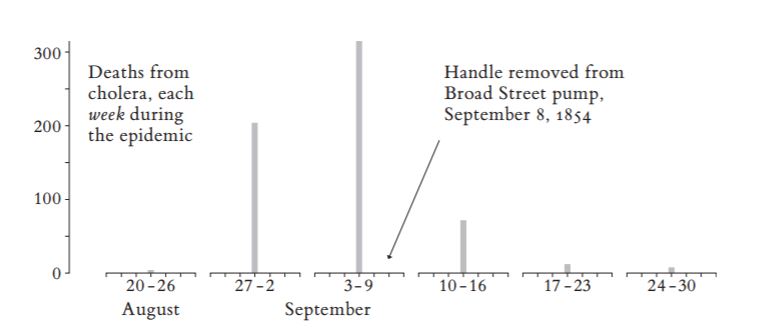
The map could identify who escaped the disease. Nobody died at the brewery which is identified in yellow on this map. It was found out that workers drank malt water instead of water from the well. The workhouse which is also identified as yellow had a handful of deaths. The workers here were drinking from the well.

Judging from his map he was able to pinpoint the source of the epidemic. It was coming from a pump.



From this graph, it is not at all clear that the removal of the handle had anything to do with the ending of the pandemic. As you can see from the graph the deaths were already in a rapid decline. Although given these ambiguities he was able to demonstrate that the disease was transmitted through water and not air.

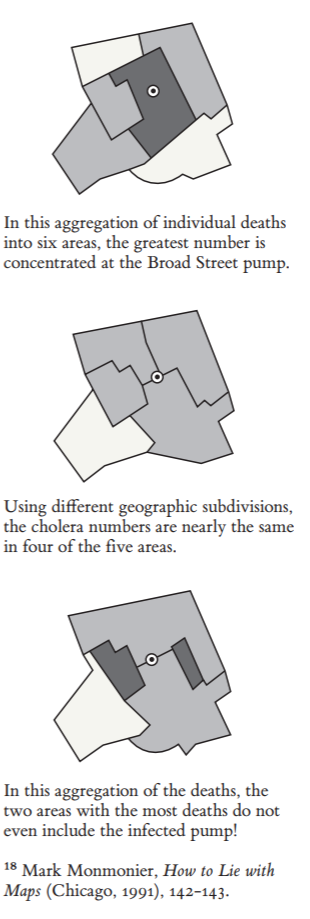
Aggregations over time would produce deceiving signals and hide relevant detail. For example, if snow showed weekly instead of daily deaths then the result would appear different. Weekly intervals would distort the output. A change of two days in weekly intervals has completely shifted the shape of the data visualisation.



* Time-series are very sensitive to the choice of intervals and endpoints.
* Aggregations can be used to reduce redundancy

Snow found a few possible errors in his graph. For instance, a patient who died in hospital, they might have come from a house that is not registered. Also, patients may have been too ill to give an account on themselves.

Snow drew a **dot map**. This marked each death.



Dot maps in this situation has advantages and disadvantages

* Death rates are not shown
* Dot maps may become cluttered with too much detail
* The deceptive effects of aggregation are avoided
* Identify individual cases

The entire area with and without cholera was thickly populated. Usually, the dot map fails to consider the number of people living in the area who are at risk of getting the disease but not in this case. Unfortunately, the dot map does not assess the densities of population around the pump well.

The cluster around the pump vanishes in the process of aggregating the data. The darker the area, the greater the number of cholera deaths.

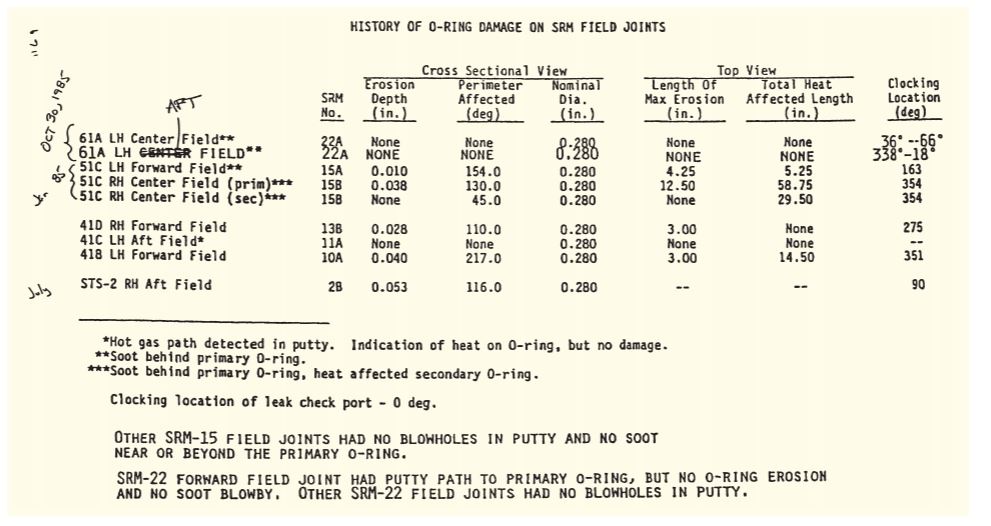
Using data visualisation techniques Dr. John Snow was able to show how Cholera was transmitted and therefore prevented.

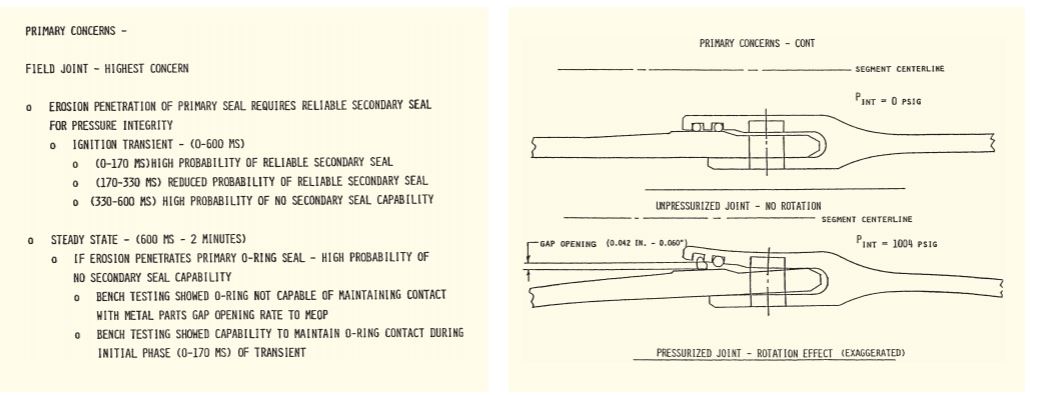
## The unfortunate decision to launch the space shuttle Challenger

This article is about a space shuttle that exploded killing seven astronauts. The shuttle exploded due to the O-ring being damaged from the cold.

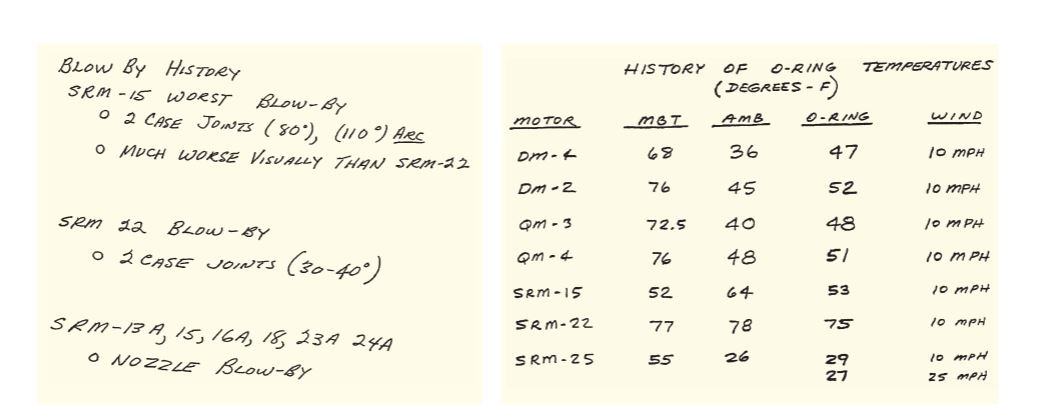
This accident has evoked the importance of data **graphs** and fitting models to data.

Engineers prepared charts on why the challenger should not be launched. However, the charts did not provide data about damage the change of temperature could do to the O-ring.



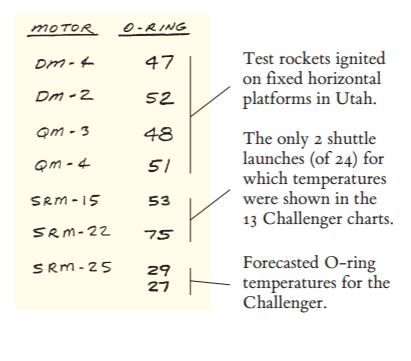


The chart above is an example, it identifies all the experienced problems with the O-rings but fail to mention temperature issues.



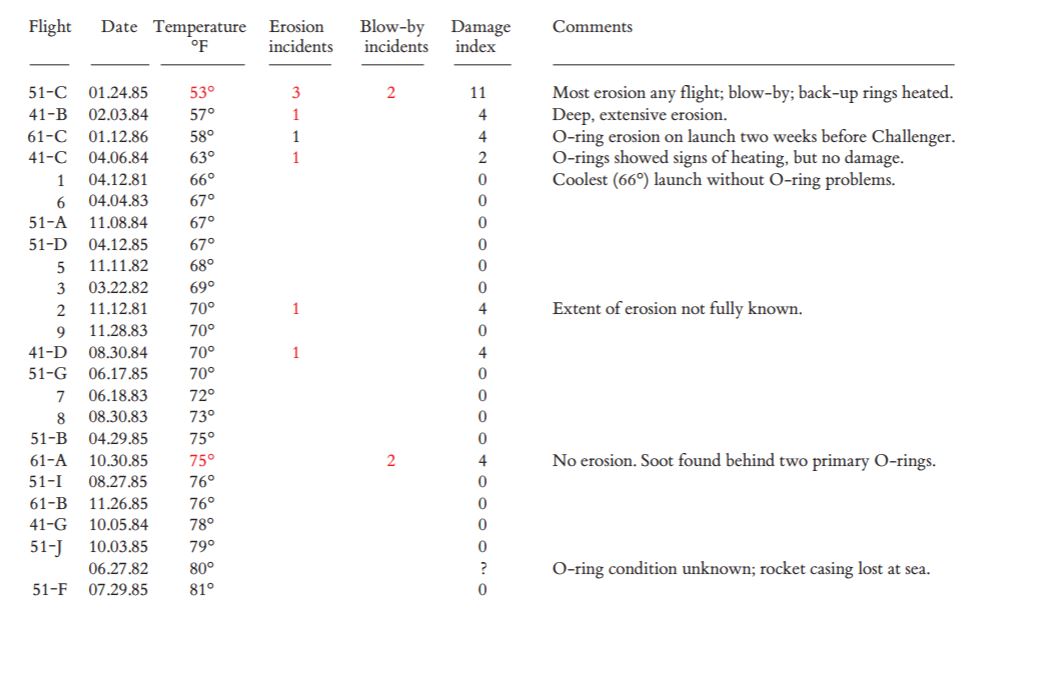
The people who decided to launch the challenger had a strong argument. The blow-by on SRM 15 was on a cool day, the blow-by on SRM 22 was on a warm day at temperature meaning they did not think the error was caused by temperature. The engineers should have investigated this further after spotting this disparity on the graph.

A lot of the temperature data is missing from the graph below. The missing data is a huge concern. 92% of the data is missing.

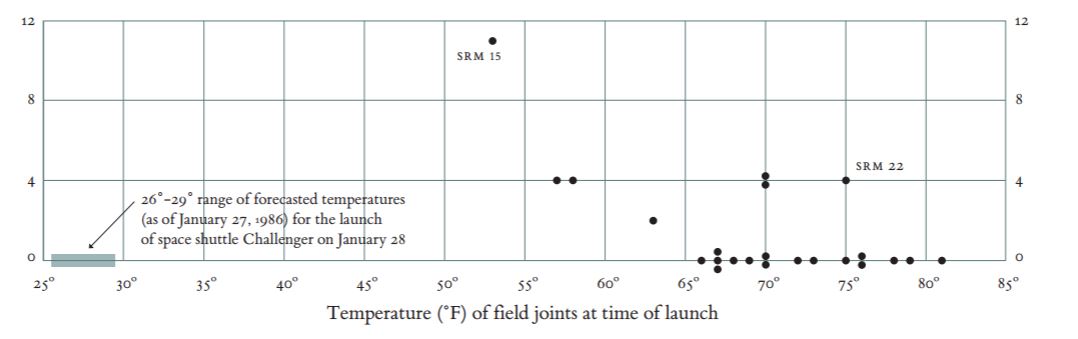


This missing data leads to an inaccurate conclusion. The shuttle could not be launched safely was the conclusion, this was rejected.

The charts failed to stop the launch but did have the right conclusion. With more concrete evidence and data, the engineers could have been persuaded to stop the launch.

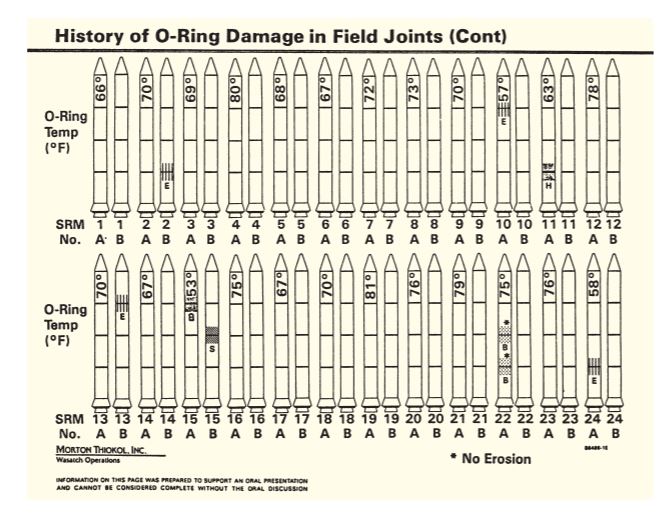


This **data matrix** shows the history of the temperature of the O-ring. This table reveals the total incidents of the O-ring on cold days compared to warm days. As we can tell from the graph more erosion incidents occur as the temperature drops.



This **scatterplot** shows the experience of all 24 launches prior to the challenger. It reveals the serious risk of launch <29º. From the graph we can see that no data was recorded for temperatures less then <50º. There was never a temperature recorded which was like the temperature at launch.

From this, we can conclude that there is a discrepancy between the tasks at hand and the images created to serve those tasks. The images failed to reveal the risks that were present. It is important to be careful about how the data is displayed. Data can be displayed in a misleading way.



This graph shows all the O-ring experiences of all 24 launches. The rockets marked with the damage code show the flights with O-ring problems. This graph has all the information needed to represent the relationship between the temperature damage. There is strong visual clutter on this graph preventing the viewer to obtain a clear concluding remark.

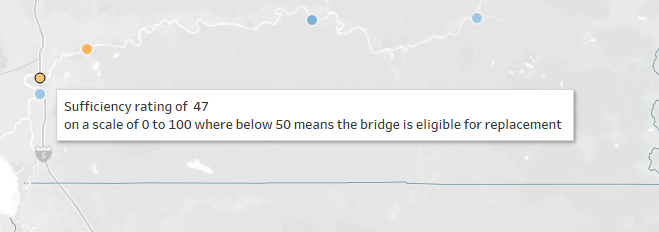
The temperature on this graph is turned sideways. The temperature is not well represented from this poor design choice.

The biggest flaw is the order of the data. The sequential order conceals the link between temperature and O-ring damage. The rockets could have been ordered by temperature which could have shown that cooler temperatures damage the O-ring. It was found that a scatter plot could have displayed the necessary information to prevent this catastrophe.

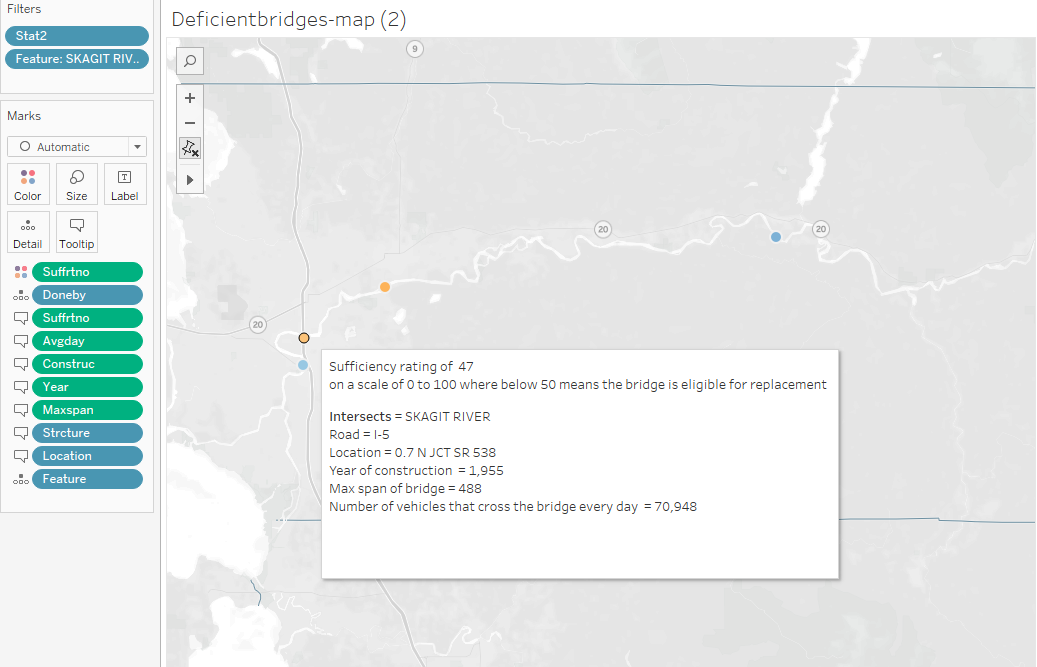
# Q2) Build a story board to show deficient bridges (slide 17) and explore the unsafe bridges (slide 20). Build the story around the causes of the collapse of the Skagit River Bridge.

## What’s the rating of the Skagit River bridge that collapsed into the river?

47

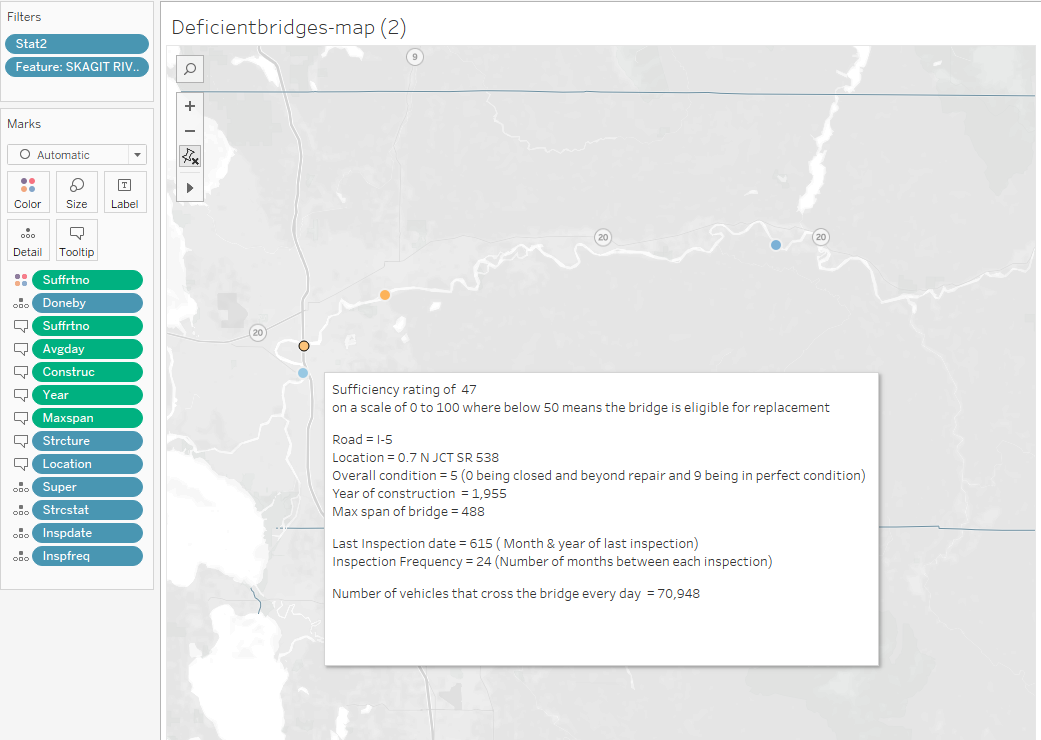


## What other distinguishing information can we obtain that might go into a story?



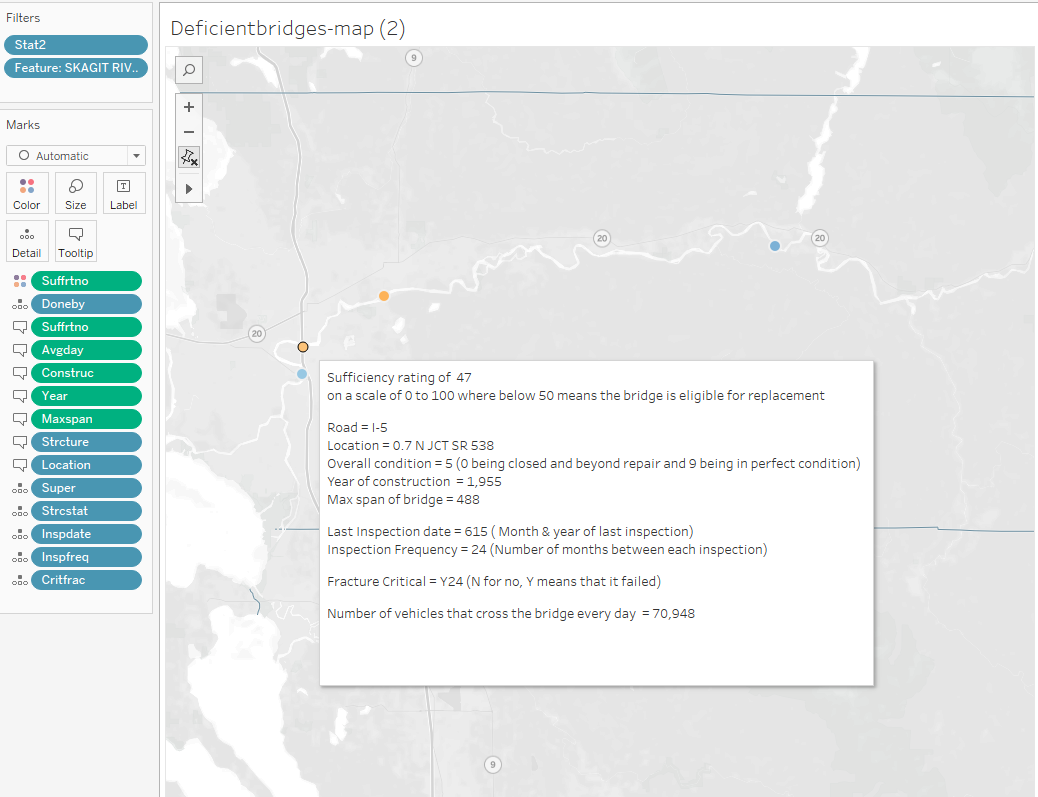
## When was the bridge last inspected and how often is it inspected?

The bridge was last inspected 6/15 and it is inspected every 24 months



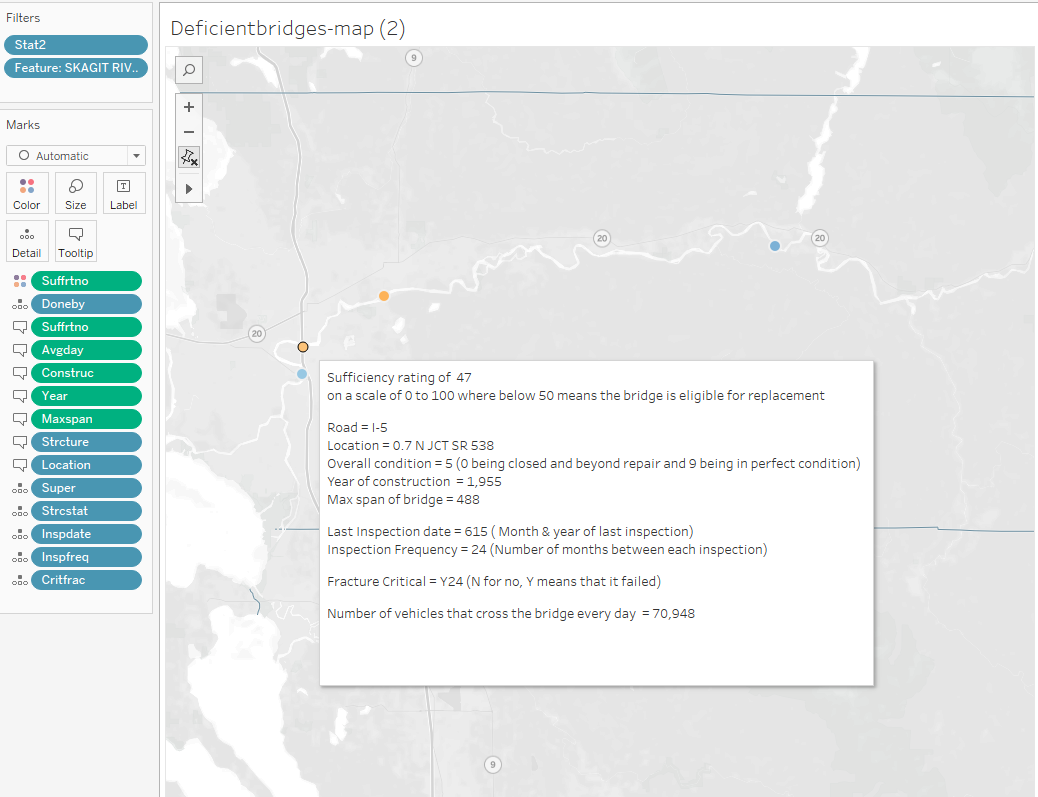
## Is it considered fracture-critical, meaning one hit to a span can knock it out?

Y24 is the result

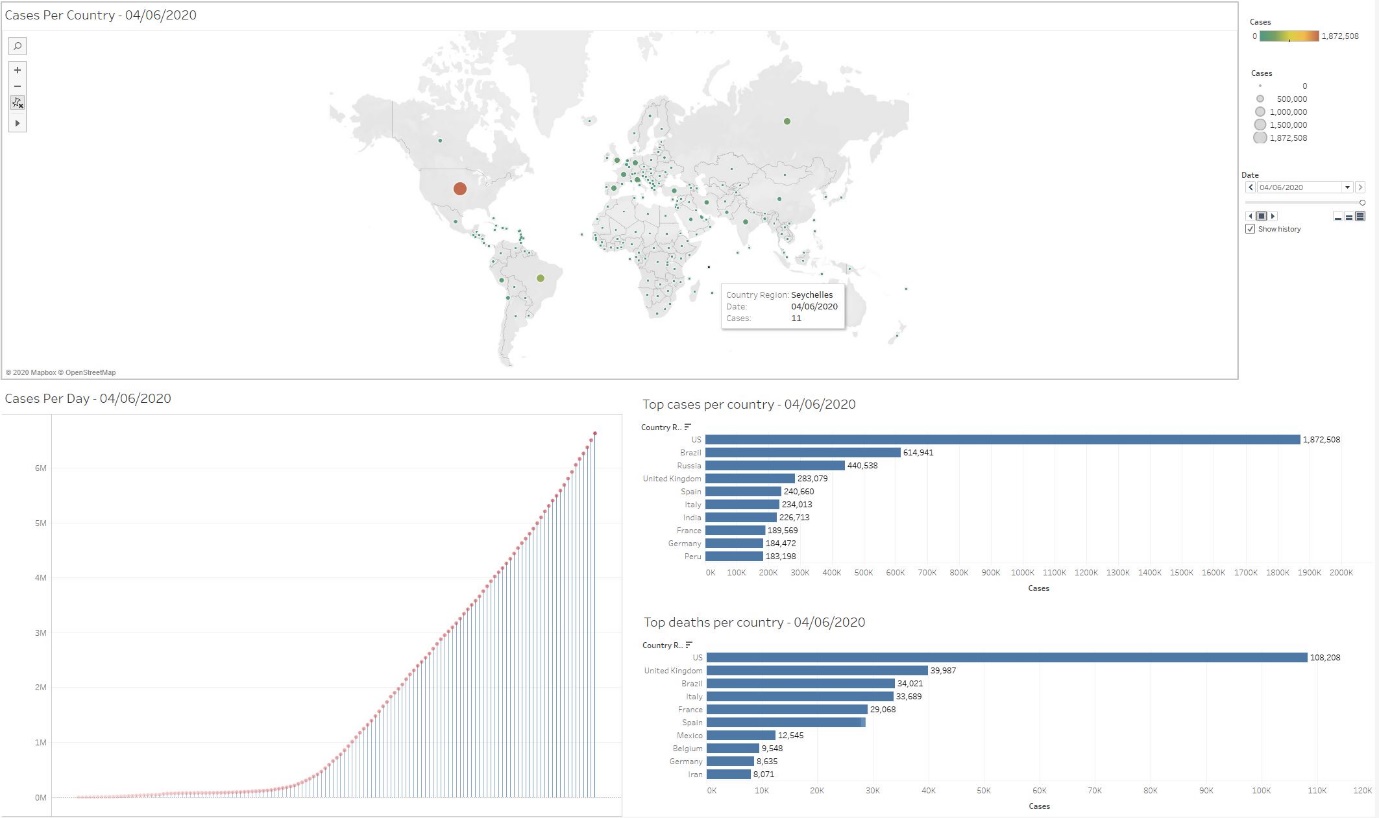


## How much average daily traffic does it get?

70,948



# Q3) Build a Tableau COVID-19 Dashboard including three visualizations telling a story related to the pandemic?



1. Map of total cases
2. Confirmed cases per day
3. Top cases and deaths per country