

# **San Diego Potholes**

## **Tracking Axle Busters in America's Finest City**

UCSD/Bootcampspot ETL Project 2

Project hosted at  
<https://sdcpothole.herokuapp.com/>

April 14, 2021

Amir Vasquez  
Jeriel Minguillan  
Tolga Caglar  
Kevin Mickey

## **Premise:**

In 2015 the City of San Diego (the City) began a concerted effort to repave and repair streets that had been neglected for decades. Potholes are reported at a near-steady state during dry weather, but pothole reports will increase during the rainy season or years when there is above average precipitation.

## **Goal:**

Discover correlations (if any) between weather and pothole formation.

Determine whether the City is seeing an annual decrease in pothole reports

Use this information to:

- Improve road repair crew scheduling

- Target historically problematic roads in the City

- Set future budgets for road repair

## **Background:**

In 2015, Mayor Faulconer initiated a program to accelerate the repaving of streets or the repair of potholes, without raising taxes to do so.<sup>1</sup> In 2016 the City began using their “Get it Done (GID) website to track pothole reports.

Using data collected by GID we were able to identify the report date, fix/closure date and location of every pothole reported to the City. GID allows the reporting party to locate the pothole by using location data on the party’s phone or tablet, or if using a non-GPS enabled device (a laptop or desktop computer) by dragging a marker to the approximate point on a map. The lat/long data can contain up to 6 digits after the decimal point, which is extremely accurate<sup>2</sup>. In reality it is much more precise than a consumer-grade GPS (7-13M)<sup>3</sup> can provide, not to mention accounting for the positional offset of the device in relation to the pothole. As a result, there is a significant chance that many of the reports are duplicates, but determining how many, and which, is beyond the scope of this project.

San Diego’s Mediterranean climate is split between a dry season that generally runs between April and October, and a wet season that runs between November and March. For official records, the water year begins on October 1<sup>st</sup> and ends on September 30<sup>th</sup> of the following year. The official rainfall for the county is measured at San Diego International Airport and has averaged just over 10” for the last 20 years.<sup>4</sup>

---

<sup>1</sup> <https://www.kpbs.org/news/2015/oct/20/san-diego-mayor-roads-repairs-no-more-taxes/>

<sup>2</sup> [http://wiki.gis.com/wiki/index.php/Decimal\\_degrees](http://wiki.gis.com/wiki/index.php/Decimal_degrees)

<sup>3</sup> <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0219890>

<sup>4</sup> <https://www.sdcwa.org/your-water/reservoirs-rainfall/rainfall/?q=/annual-rainfall-lindbergh-field>

## Data Sources

City Council district map geojson: <https://data.sandiego.gov/datasets/city-council-districts/>

Pothole reports via City of San Diego *Get it Done* website: <https://www.sandiego.gov/get-it-done>

Weather: 5 year data download from NOAA in CSV format<sup>5</sup> (Submitted via web interface request)

We obtained weather data from January 2016 to March 2021.

The pothole reporting system went live in late May 2016 and continues through the present, but we cut it off on March 31, 2021.

## Methods/Tools

Heroku hosted PostgreSQL

PGAdmin

Flask

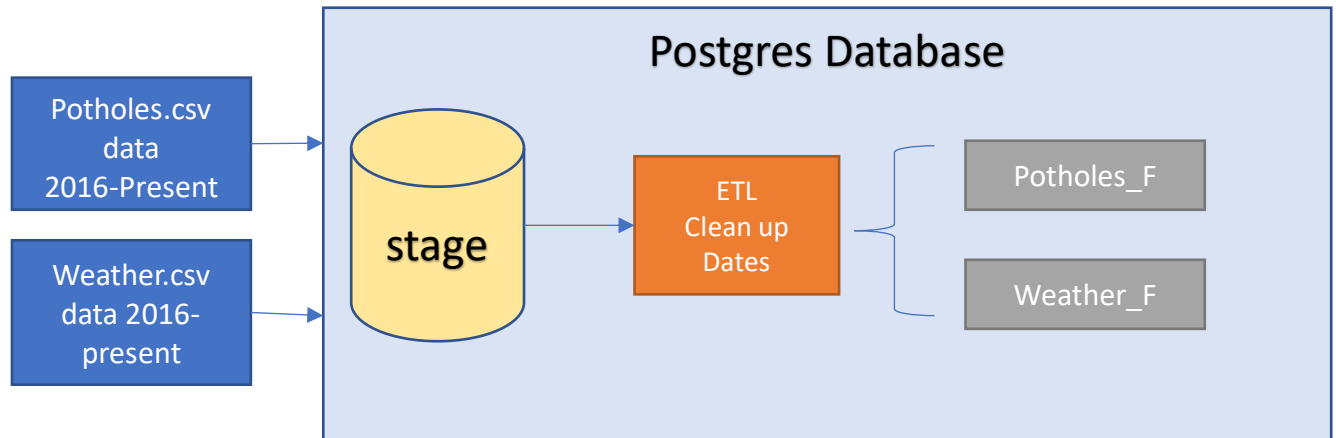
Excel (to validate/examine CSV files)

JSON Formatter/Validator <https://jsonformatter.curiousconcept.com>

---

<sup>5</sup> <https://www.ncdc.noaa.gov/cdo-web/search>

## Data Schema



### Stage area:

Sets all elements from the files as is in TEXT columns

### ETL

Store Procedure cleans and standardizes dates for use with Date Dimension

## Summarizing Data

Because Heroku has a 10k record limit on their free tier, we used summary tables to condense our approximately 40K weather observations and 80K pothole reports.

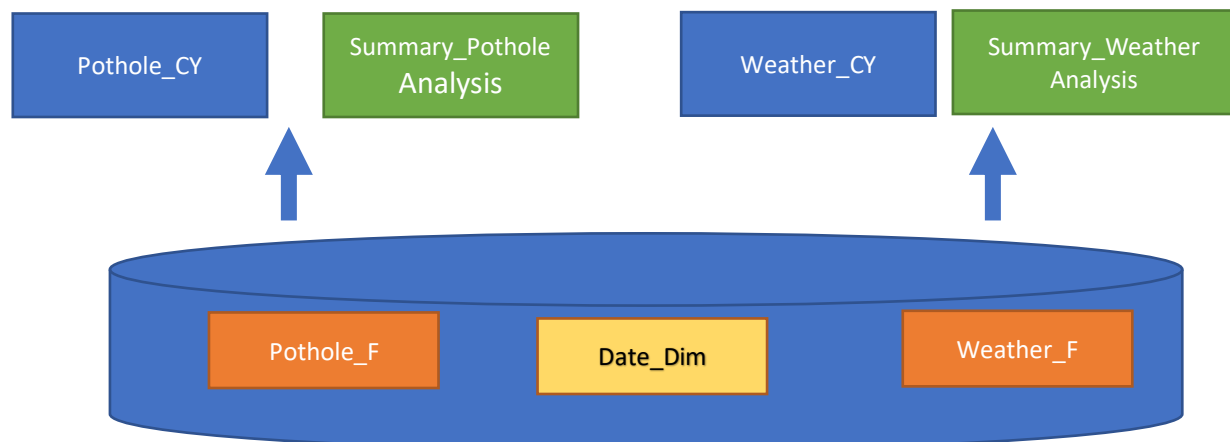


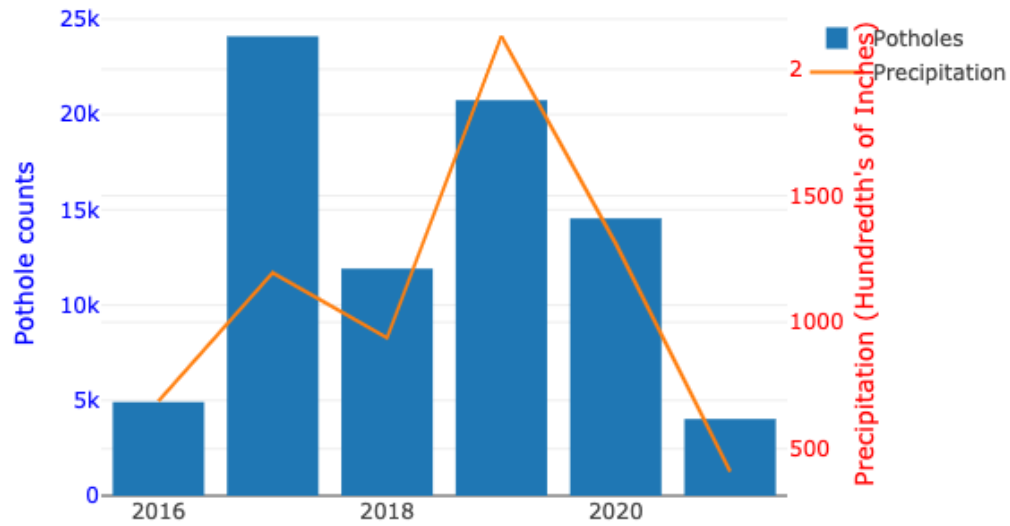
Table: public.summary_potholeAnalysis			
Field	Type	Size	Example Value
pksummaryid	integer		1
year_actual	integer		2016
quarter_name	character	varying(10)	3Q-2016
month_actual	integer		8
month_name	character	varying(5)	Aug
status	character	varying(40)	Closed
cnt_new	integer		0
cnt_inprogress	integer		0
cnt_closed	integer		433
cnt_referred	integer		0
total_cnt	integer		433
<b>Description</b>			
This table contains summary level metrics of potholes that are New, in progress, closed, referred grouped at the month, quarter and year level.			

## Charts:

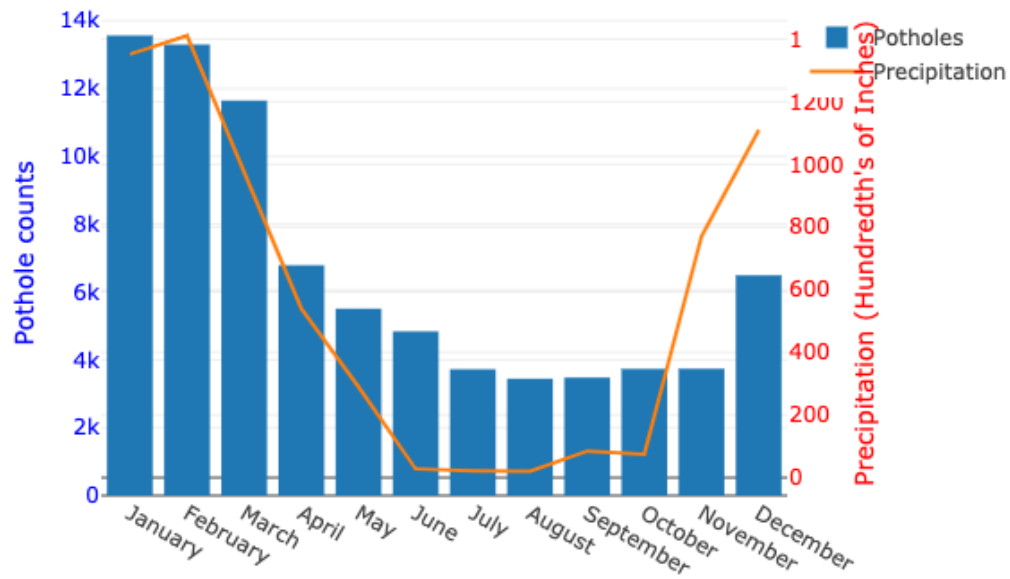
In the chart below there appears to be a correlation between pothole formation and rainfall. 2016 was the inaugural year for GID, so the numbers are probably much lower than they would've been if the program had been running in 2014 or 2015. It's a safe assumption 2017 contains reports for potholes formed in 2016, but were not reported due to people not knowing about GID

The second chart demonstrates the relationship between monthly precipitation and pothole formation. There is a clear increase in pothole reports as precipitation increases.

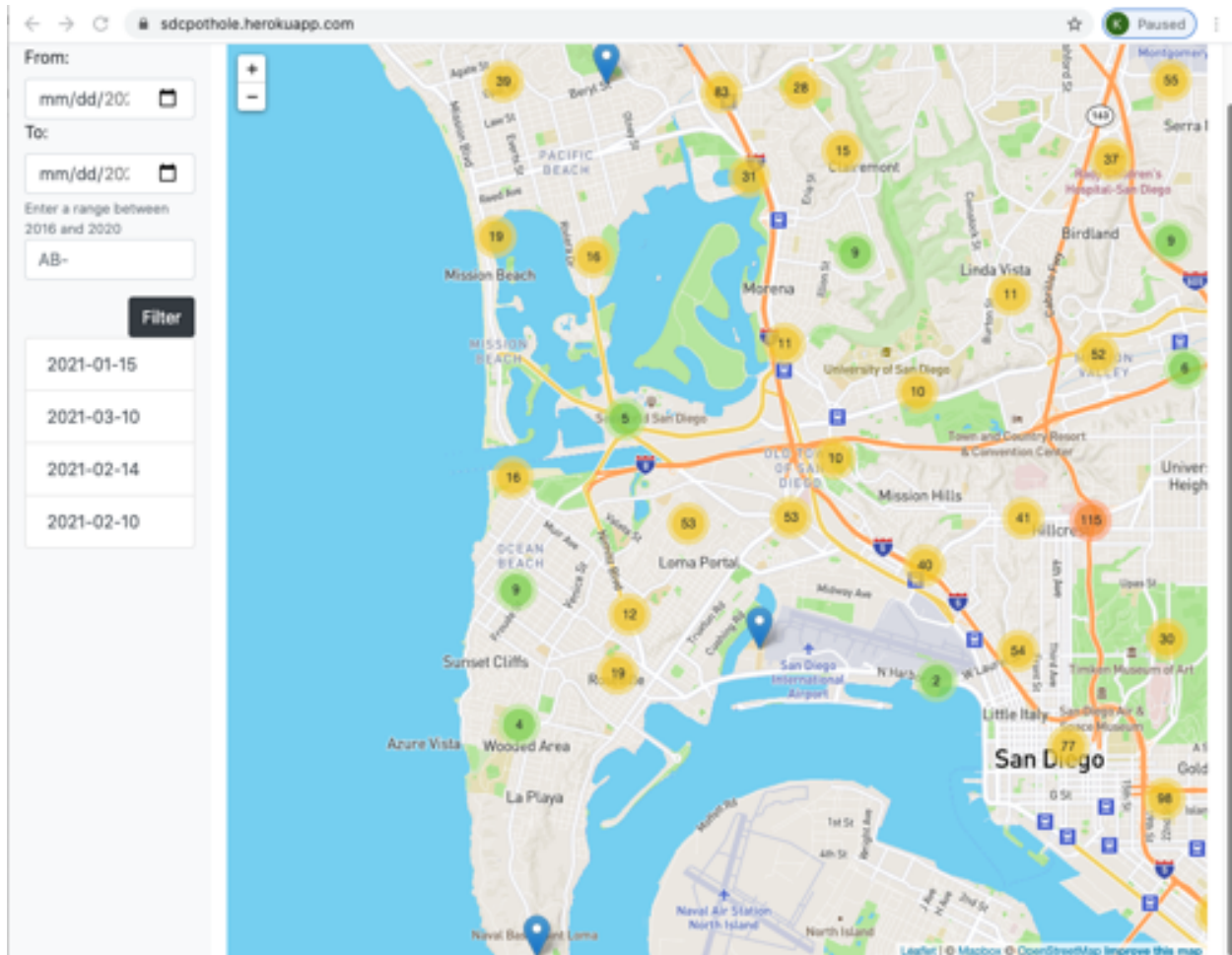
### Potholes per year



### Potholes per month



## Map(s):



### Key

- 1-4 days = green
- 5-8 days = yellow
- 9 or more days = red

## Issues

- 500 server issues heroku concurrency – solved with Renan assist during office hours
- index.html not generating map
- **Connecting all the moving parts without breaking others. We had a successful deployment at approximately 2300 13 April 2021, but we changed something an hour later and broke it badly.**
- Site partially restored 1347 14 April. Potholes work, but city council overlay broken
- City Council overlay restored 1730 14 April