ENSF 592 Final Project

Authors: Patrick Pickard, Joshua Posyluzny

Presentation Date: August 13, 2020

Project Topics

- Area of Interest
- Grid The City
- Master DataFrame
- Data Visualization
- Conclusion

West bound:	-114.315796
East bound:	-113.859905
North bound:	51.212425
South bound:	50.842822

Area of Interest

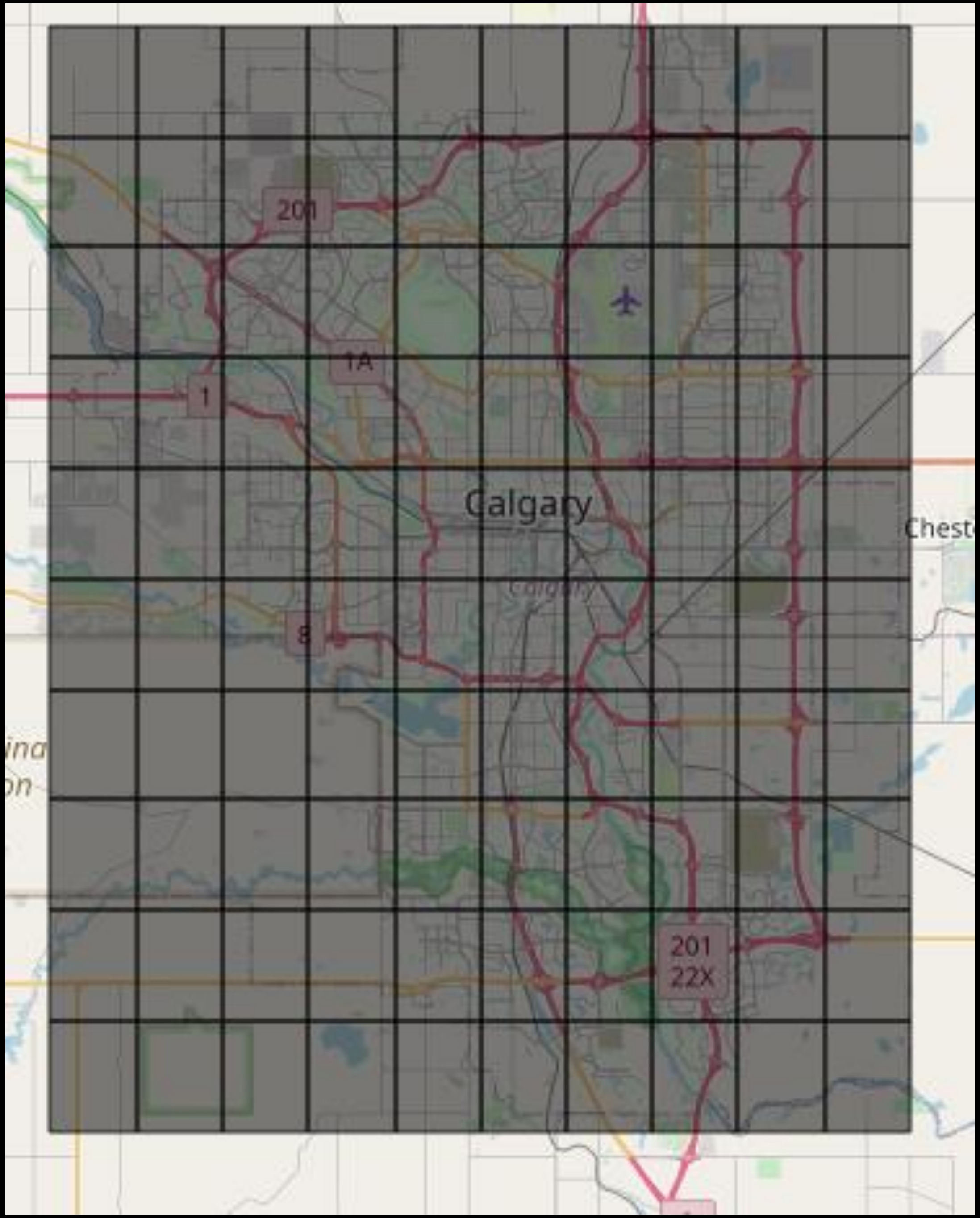
The goal of this project is to analyze various traffic data inside the following orange square section, which defines the maximum boundaries of the city of Calgary, and attempt to draw correlations between the data and the number of incidents that occurred inside this area in 2018.

Cleaning the Datasets

- All csv files were read into panda dataFrames. Any character encoding errors were stripped out and replaced. Columns that were not pertinent or useful to our analysis were dropped and the remaining columns were ordered . Date and time data were formatted to correctly. The columns containing the data we planned to use were sorted.
- Any hourly datasets (weather visibility data), were resampled for daily averages and appended to existing weather dataFrame.
- The multilinestring coordinates for locations were formatted from lon, lat to lat, lon.

```
speed_limit_df = pd.read_csv("Speed_Limits.csv")

# cleaning the speed_limit data
speed_limit_df['multiline'] = speed_limit_df['multiline'].str.replace(r'[MULTILINESTRING]','').str.replace(r'\(','').str.replace(r'\)
','').str.replace(r'\)','').str.replace("-", "-").str.strip()
speed_limit_df.dropna(subset = ["SPEED"], inplace=True)
```



Grid The City

The city grid was built using the pd.cut function from Pandas to build 10 equal size bins for both latitude and longitude using the max and min of each. The bins were then used to create the grid breakdown inside the DataFrame and the GeoJson file that was used to generate the Choropleth maps.

	lat_high	lat_low	long_high	long_low	incident_count	camera_count	average_speed_limit	average_traffic_vol	signals_count	signs_count
1	51.212425	51.175465	-114.316252	-114.270207	0	0	0	0	0	0
2	51.212425	51.175465	-114.270207	-114.224618	0	0	75	0	0	18
3	51.212425	51.175465	-114.224618	-114.179029	0	0	65	3000	0	84
4	51.212425	51.175465	-114.179029	-114.133440	11	0	64	4500	1	845
5	51.212425	51.175465	-114.133440	-114.087850	2	0	60	4750	2	644
...	...	...	...	...	...	...	...	...	...	...
96	50.879782	50.842452	-114.087850	-114.042261	7	1	60	10000	1	303
97	50.879782	50.842452	-114.042261	-113.996672	17	1	73.3333	9500	5	1487
98	50.879782	50.842452	-113.996672	-113.951083	13	0	80	16000	7	756
99	50.879782	50.842452	-113.951083	-113.905494	0	0	0	0	0	27
100	50.879782	50.842452	-113.905494	-113.859905	0	0	0	0	0	5

Master DataFrame

Our Master DataFrame is organized as follows:

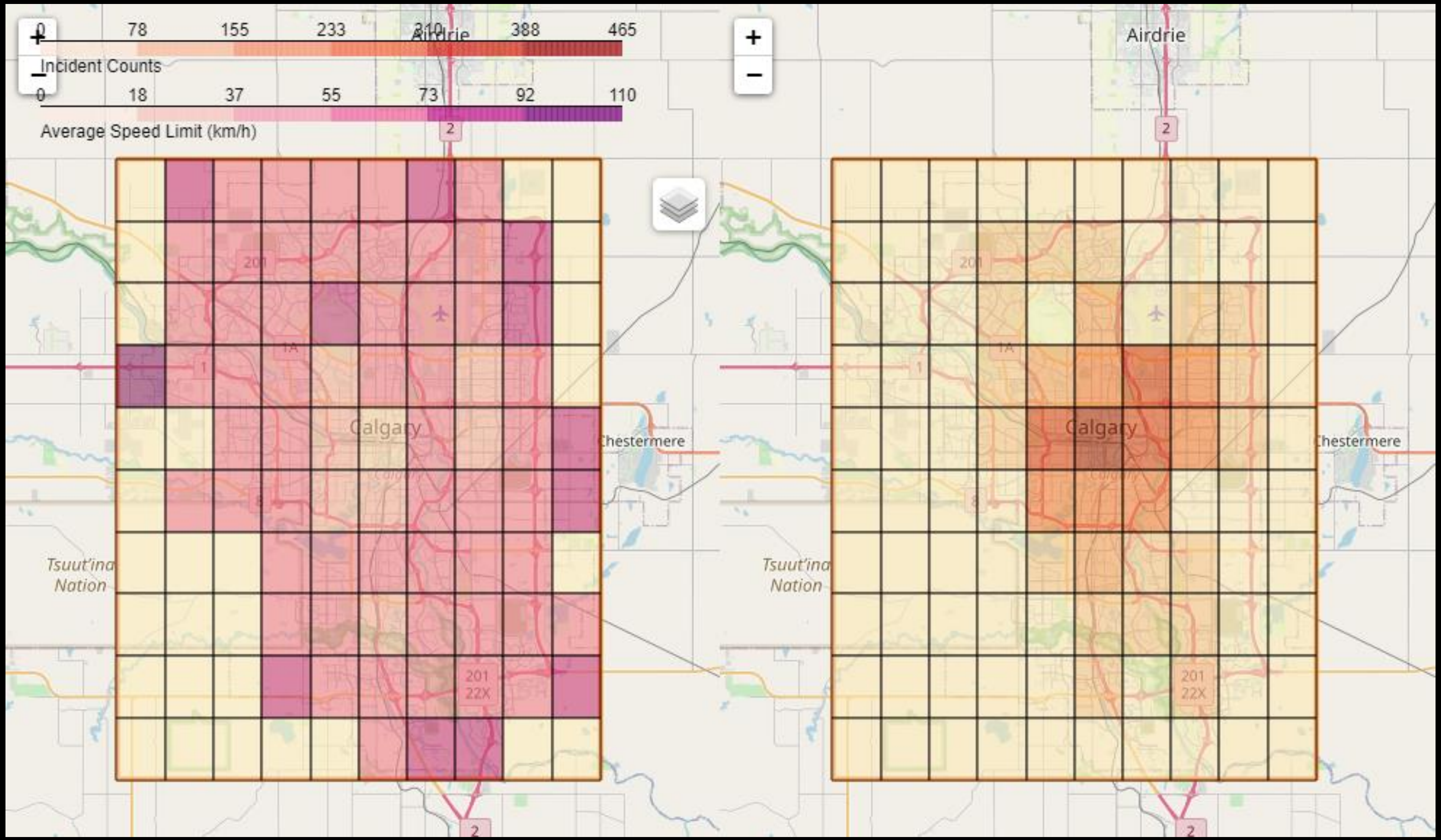
- Index refers to the grid number. Grids start at the top left of the map, and go left to right, up to down
- lat_high, lat_low, long_high, and long_low define the boundaries of each grid
- Data for the remaining columns/features are found using loops to traverse the cleaned data to find how many of each feature occur in each grid, and the counts/averages for each feature are displayed for the grid

Data Visualization

The following slides will show visualization for the analyzed data in the forms of maps and plots.

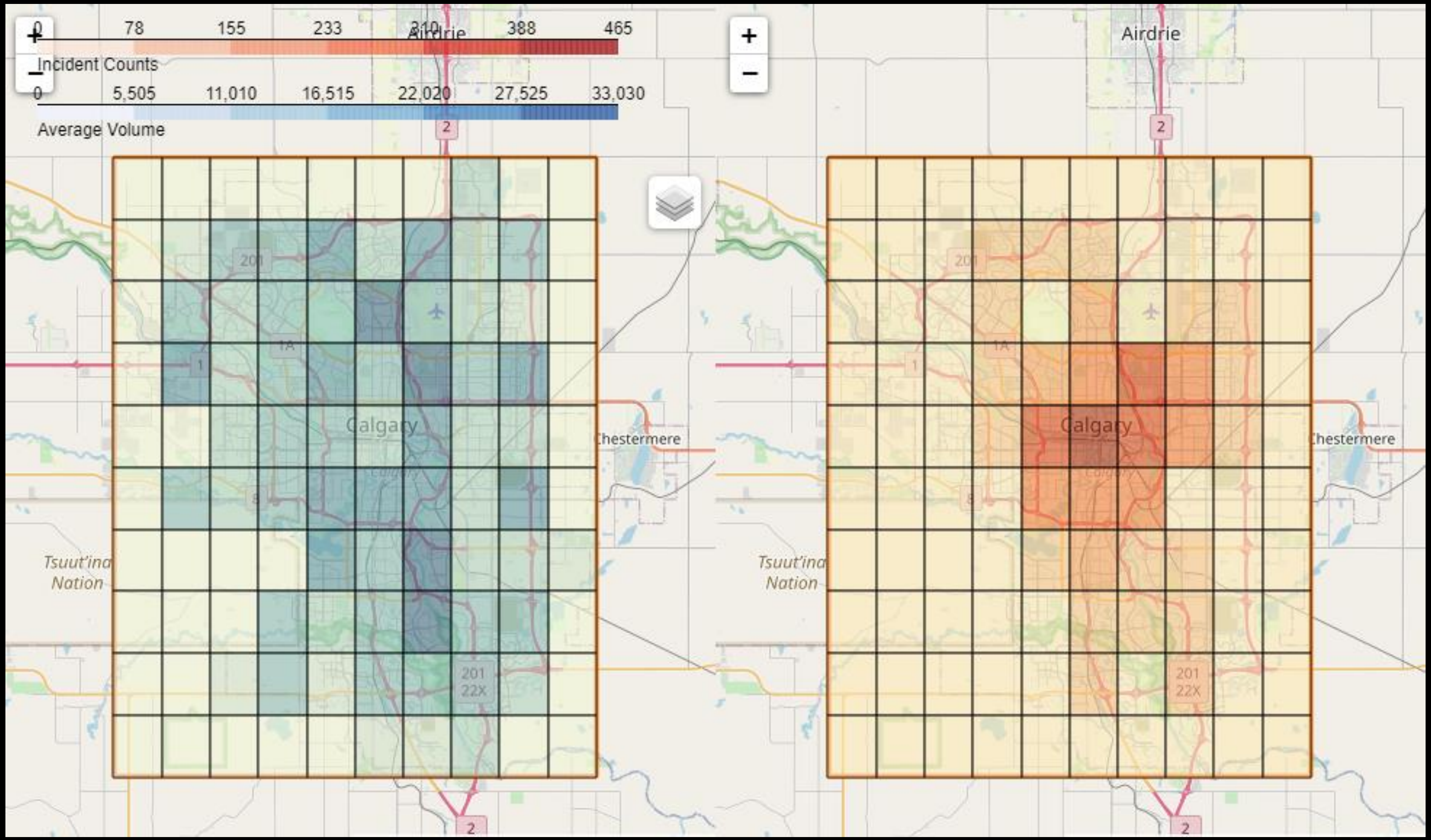
Traffic Incidents vs Average Speed Limits

These data show the lowest correlation.




```
The Spearman correlation of incidents and average speed limits is:  0.35497197270724873
```

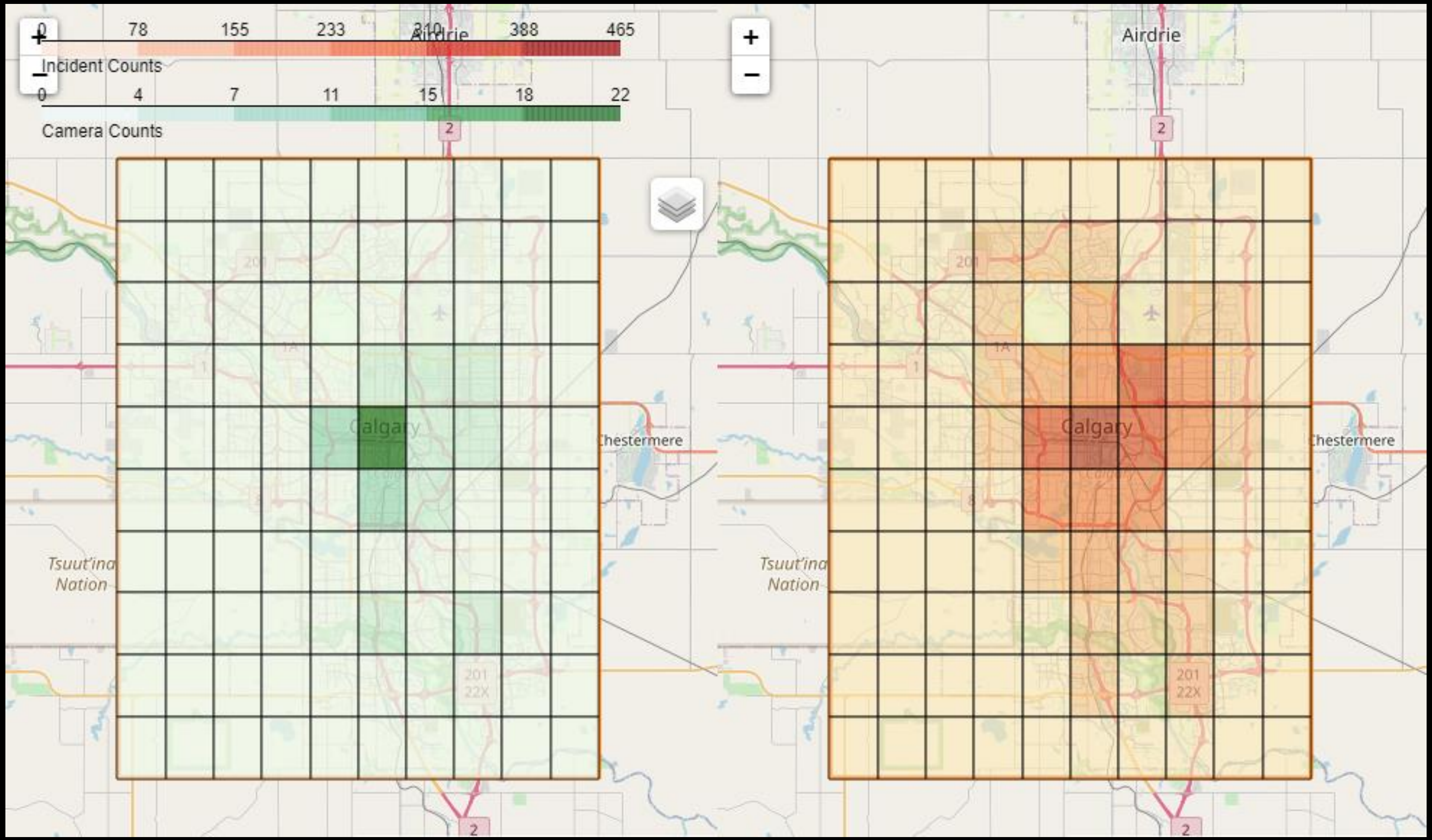
Traffic Incidents vs Average Traffic Volumes



The Spearman correlation of incidents and average traffic volumes is: 0.8426682519031534

These data show very strong correlation.

Traffic Incidents vs Traffic Cameras

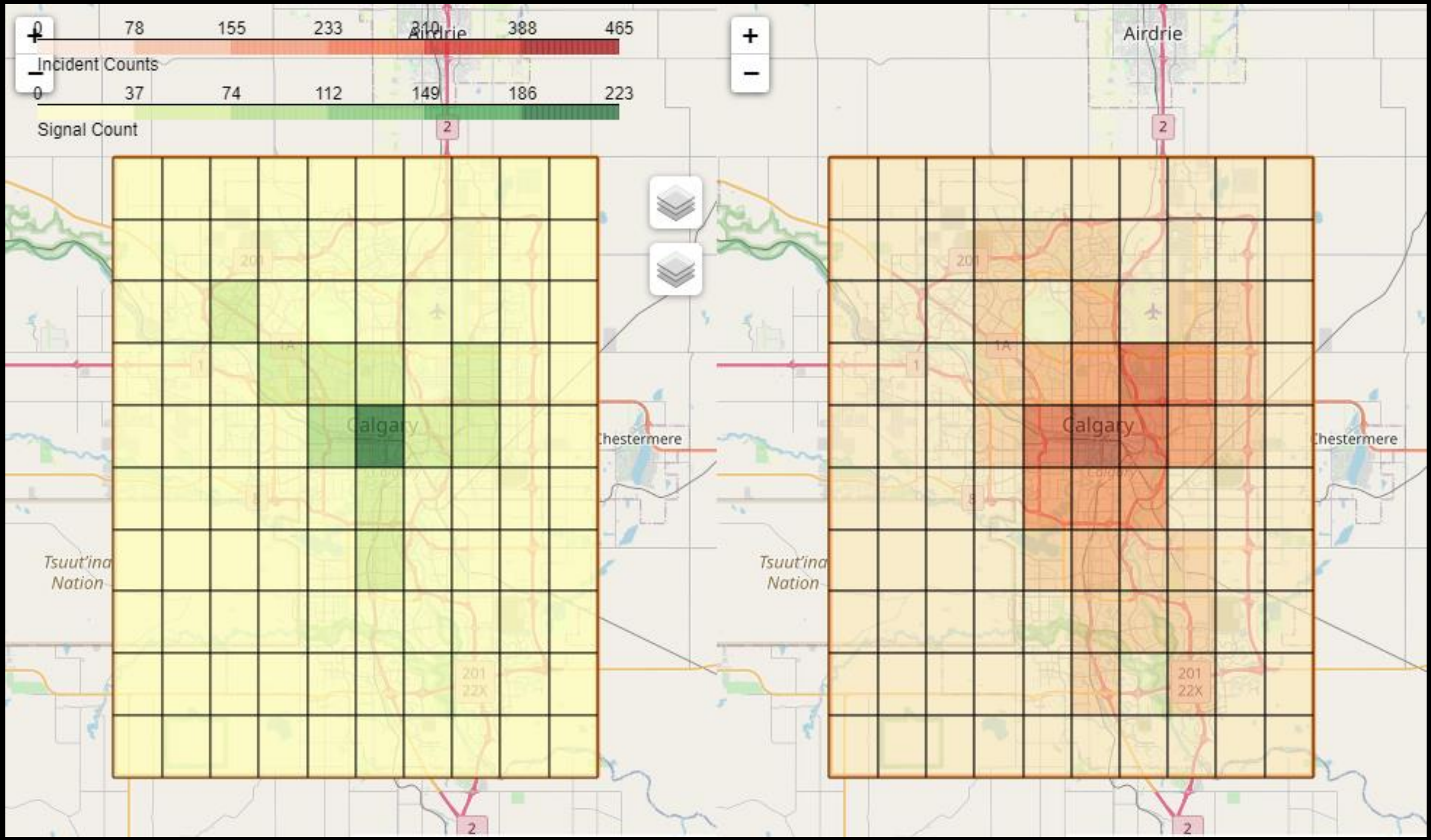


The Spearman correlation of incidents and cameras is: 0.8006178167759678

These data show very strong correlation.

Traffic Incidents vs Traffic Signals

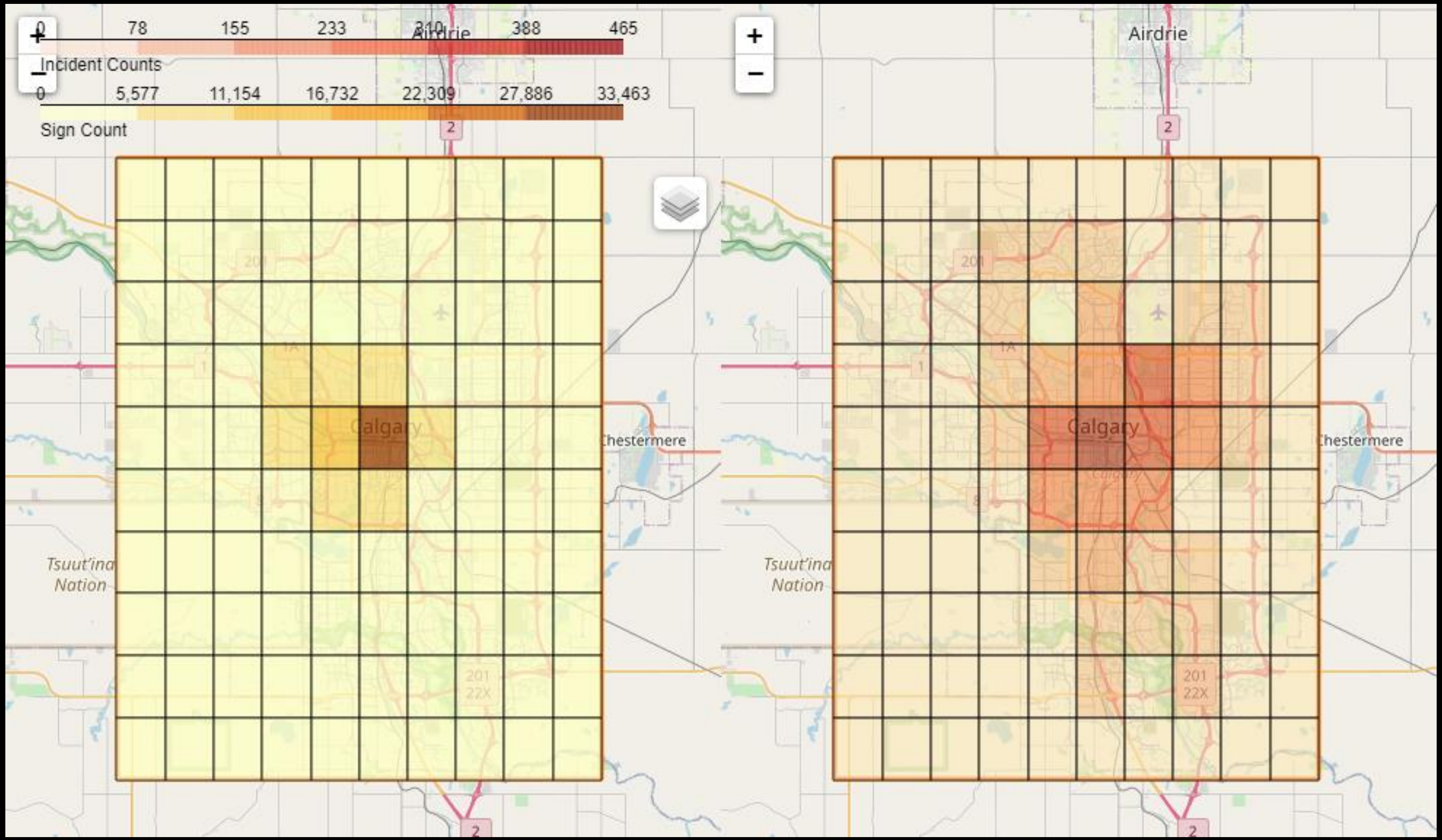
These data show the strongest correlation.



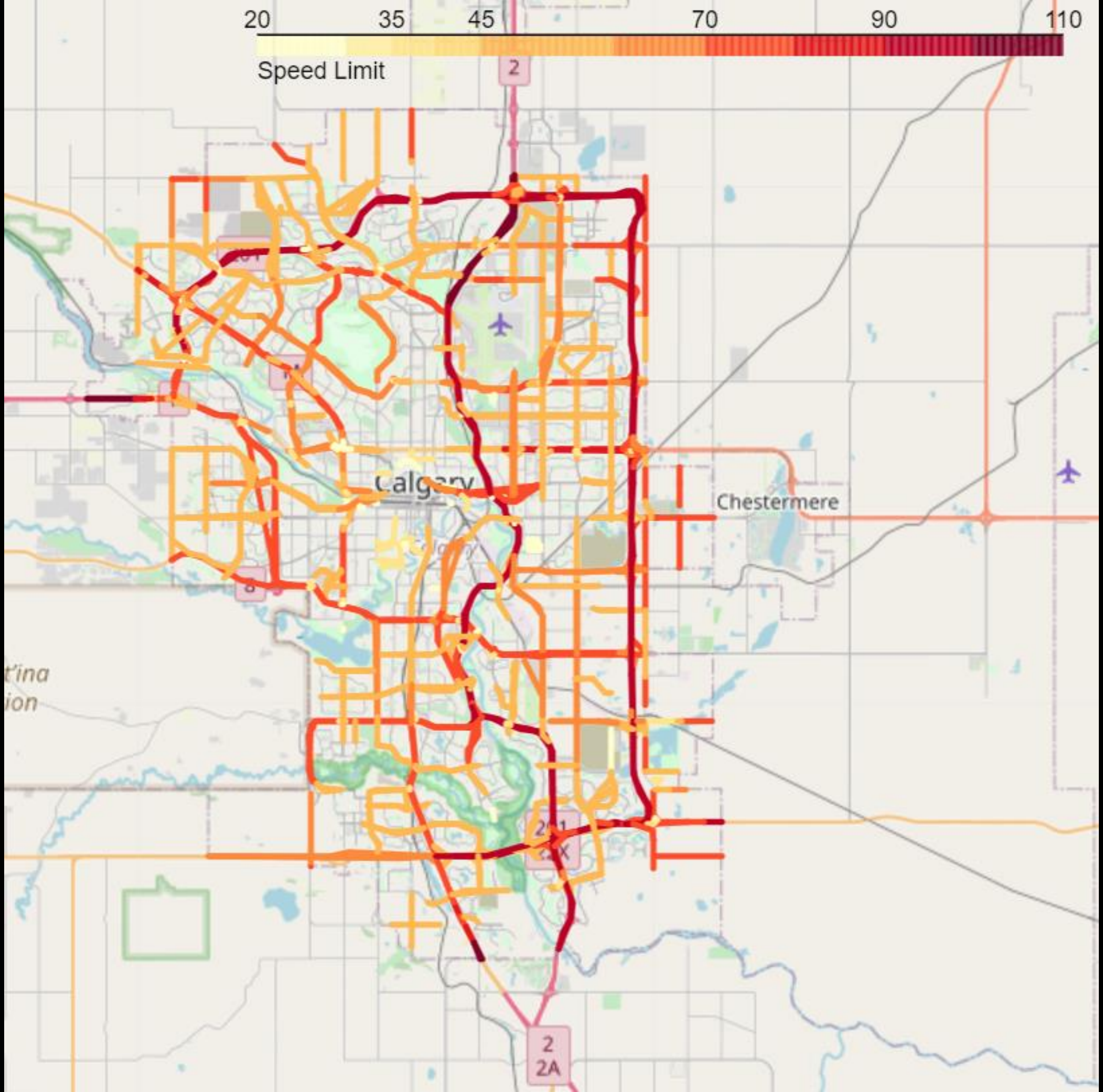

```
The Spearman correlation of incidents and traffic signals is:  0.9405777966555693
```

Traffic Incidents vs Traffic Signs

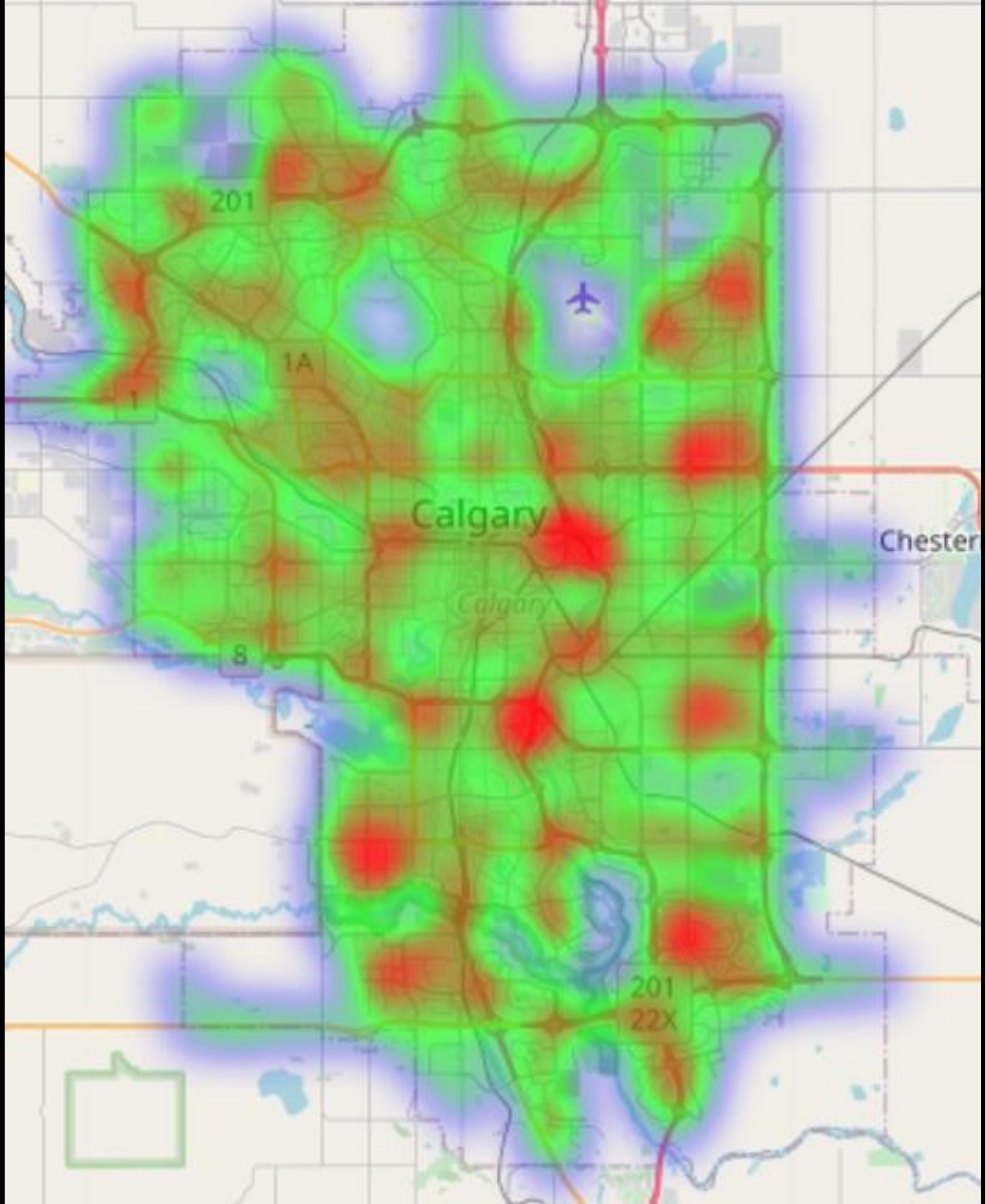
These data show very strong correlation.



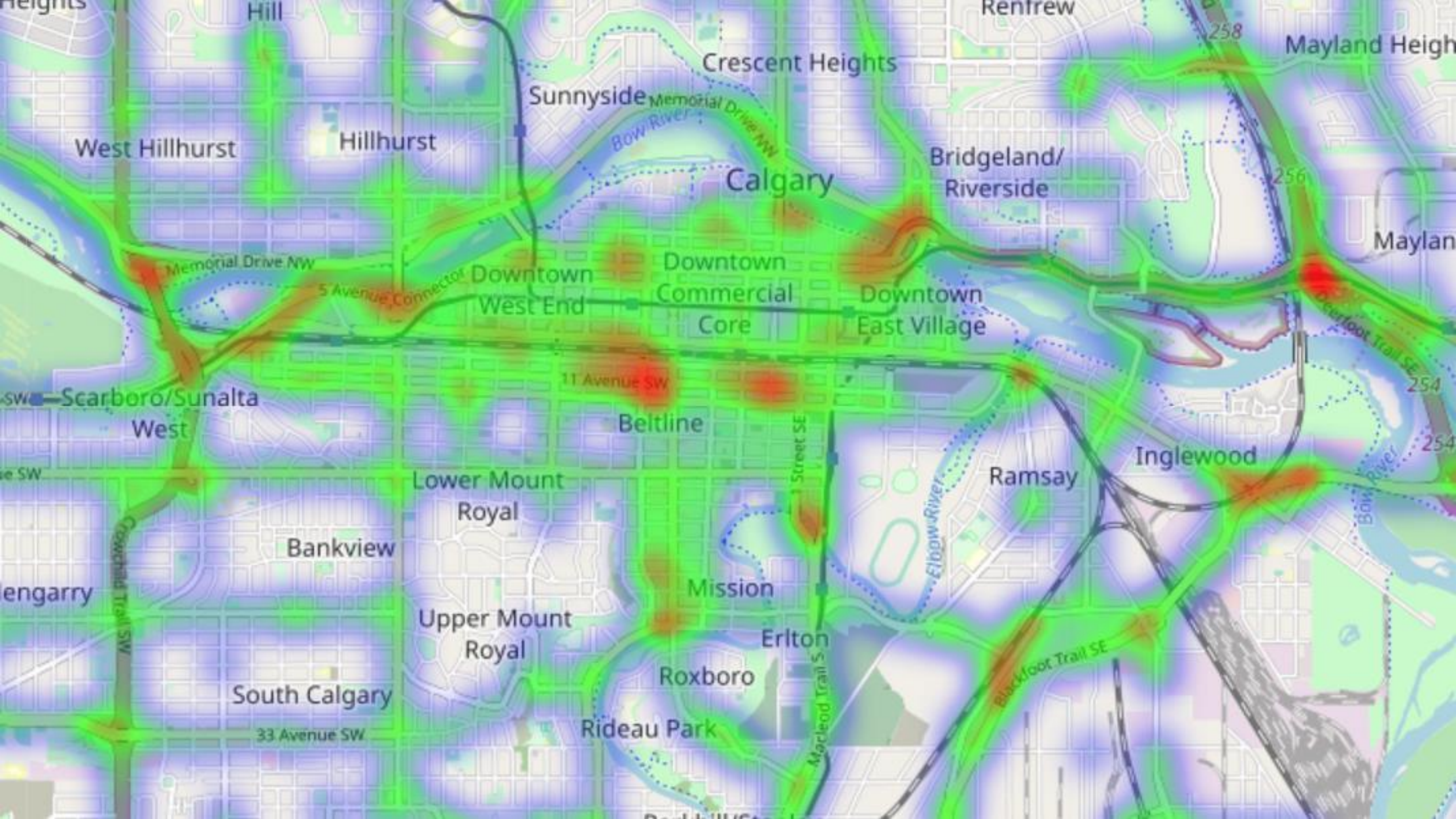
The Spearman correlation of incidents and traffic signs is: 0.929009683943728

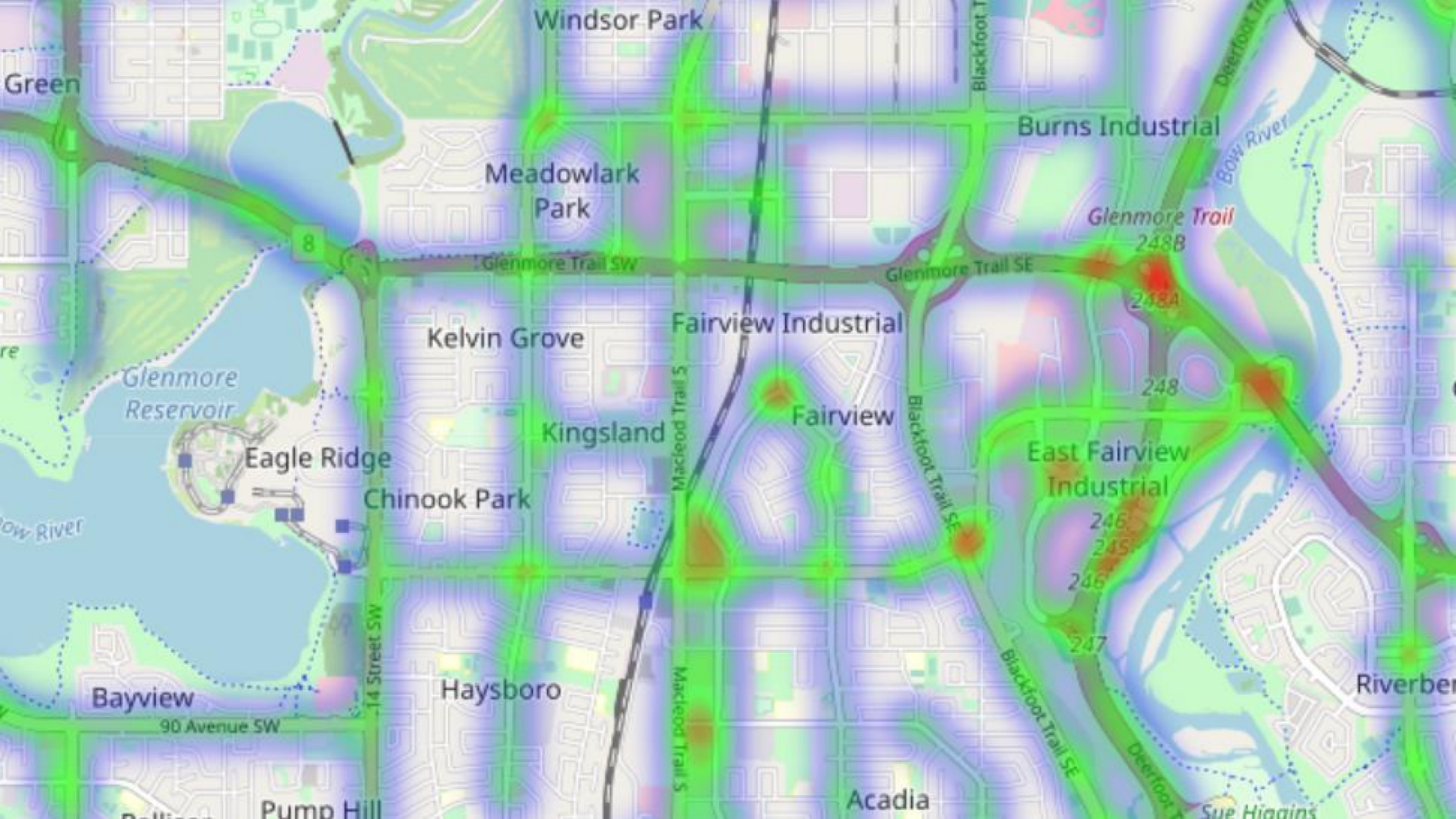


Road Speed Limits

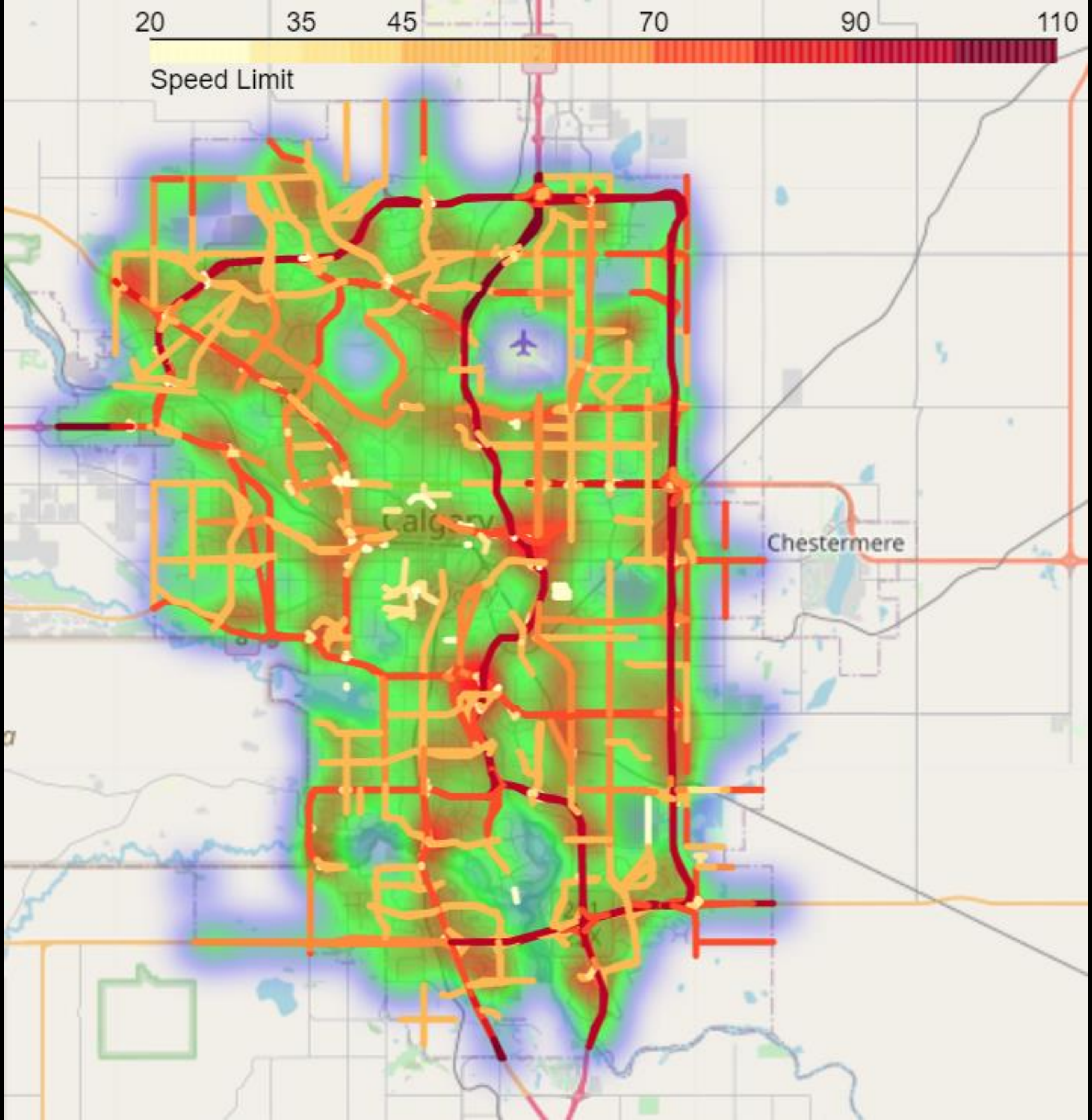


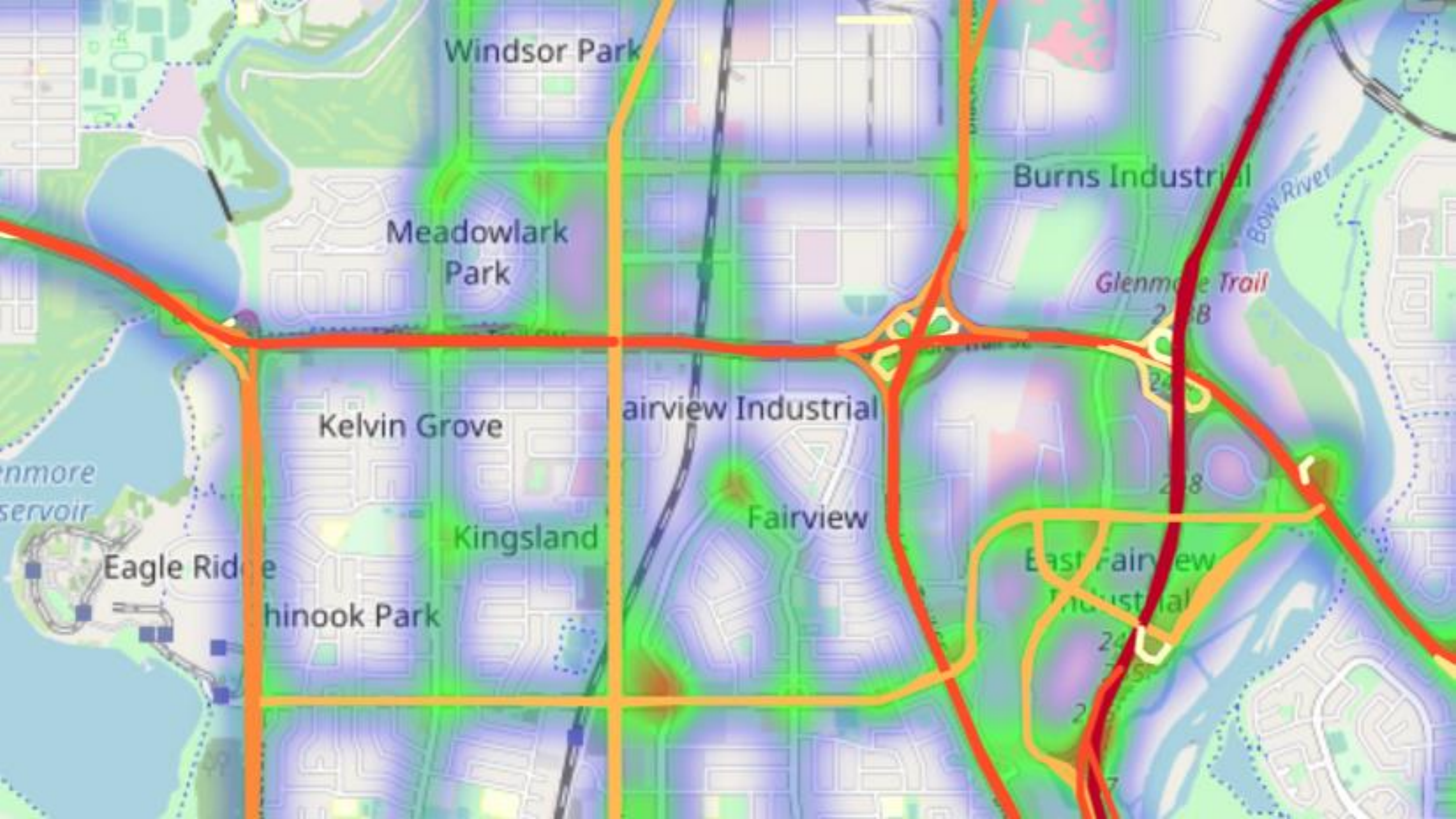
Heatmap by Traffic Volumes

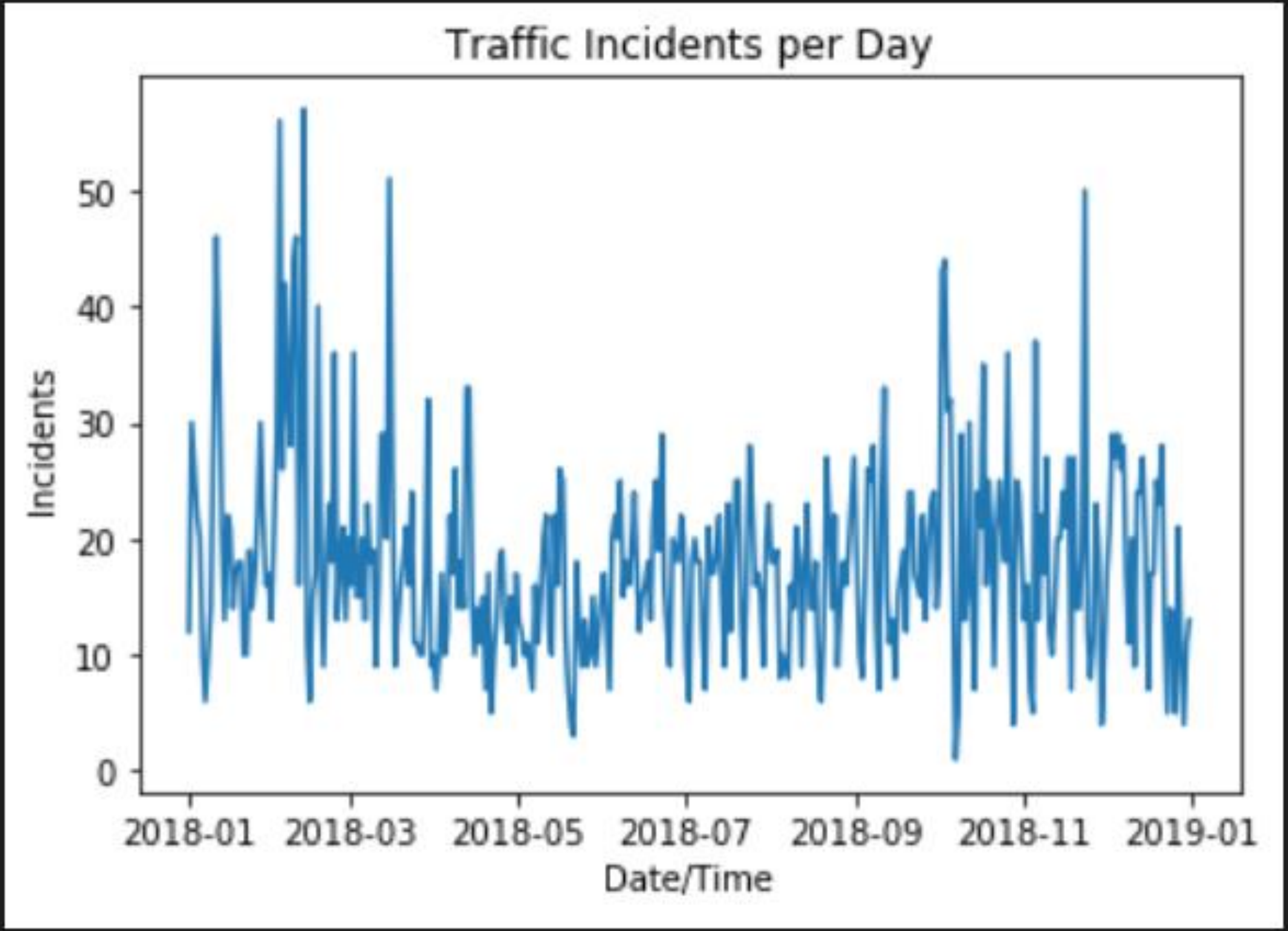

Heights
Hill
Renfrew
Mayland Heigh
Crescent Heights
Sunnyside
Memorial Drive NW
Bow River
West Hillhurst
Hillhurst
Calgary
Bridgeland/
Riverside
Maylan
Memorial Drive NW
5 Avenue Connector
Downtown
West End
Downtown
Commercial
Core
Downtown
East Village
Deerfoot Trail SE
11 Avenue SW
Scarboro/Sunalta
West
Beltline
1 Street SE
Inglewood
Ramsay
Lower Mount
Royal
Bankview
Crowchild Trail SW
Elbow River
Bow River
Mission
lengarry
Upper Mount
Royal
Erlton
Blackfoot Trail SE
Roxboro
Macleod Trail S
South Calgary
33 Avenue SW
Rideau Park
258
256
254
254


Windsor Park
Green
Burns Industrial
Bow River
Meadowlark
Park
Glenmore Trail
248B
8
Glenmore Trail SW
Glenmore Trail SE
248A
Fairview Industrial
Kelvin Grove
Glenmore
Reservoir
248
Fairview
Kingsland
Macleod Trail S
Blackfoot Trail SE
Eagle Ridge
East Fairview
Industrial
Chinook Park
Bow River
246
245
246
247
14 Street SW
Haysboro
Bayview
90 Avenue SW
Riverbe
Macleod Trail S
Blackfoot Trail SE
Deerfoot Trail
Acadia
Pump Hill
Sue Higgins

Traffic Volume Heatmap & Road Speed Limits

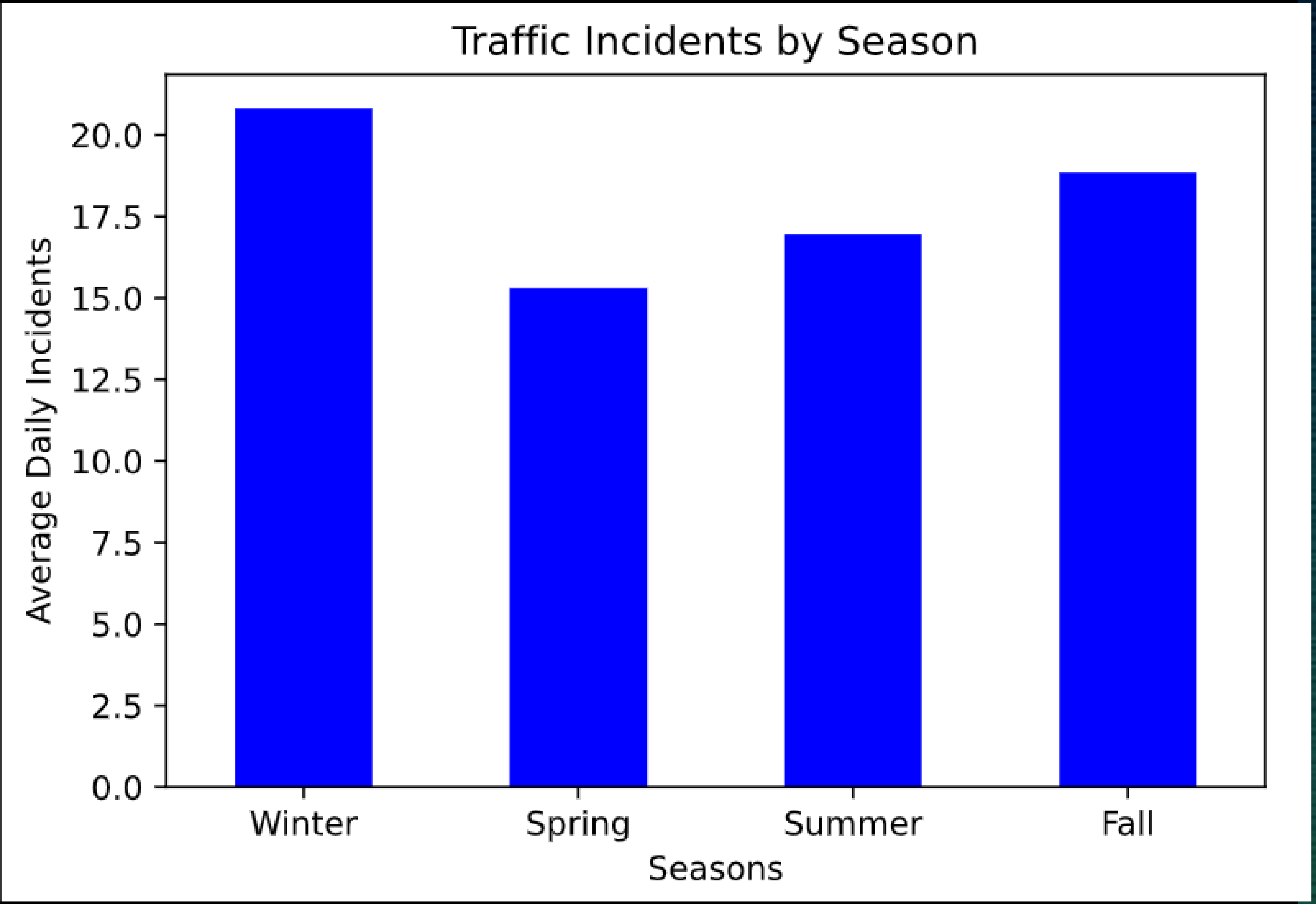



Windsor Park
Meadowlark
Park
Bow River
Kelvin Grove
Fairview
Kingsland



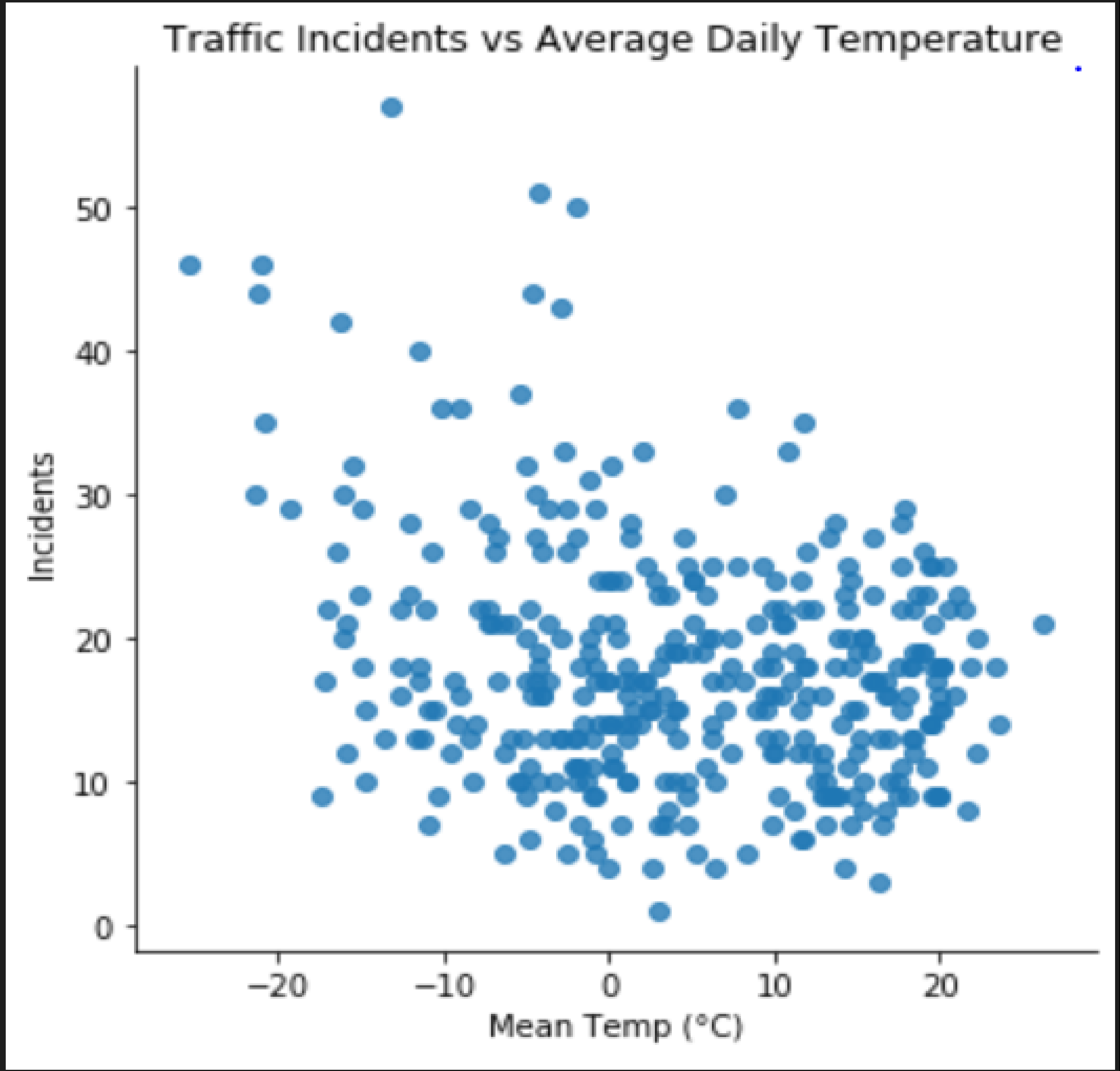
Hard Data Visualization – Plots

- There is very weak correlation between seasons and incident volumes (see next slide).
- The spearman coefficient of correlation for each of these graphs can be seen below each plot.



- The only moderate conclusion that can be drawn from this graph is that there are, on average, more incidents in the fall and winter months, October – March, compared to the spring and summer months, April – September.

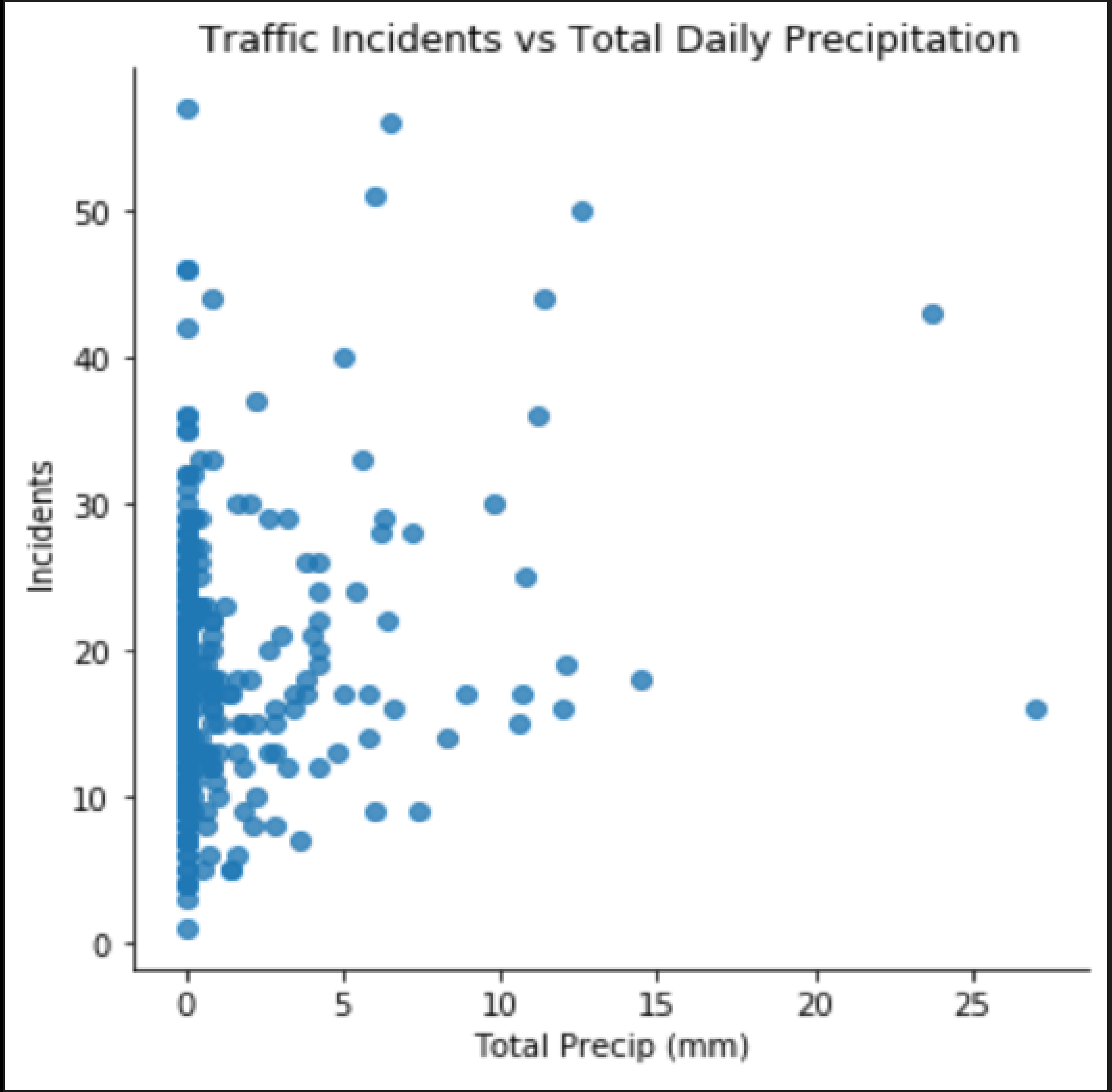
```
The Spearman correlation of incidents for by season_bins is:  -0.1238948234693196
```



Hard Data Visualization – Plots

- There is no significant correlation between the average daily temperature and the number of traffic incidents. (Very slightly more at lower mean temperatures, but not statistically signifcant)

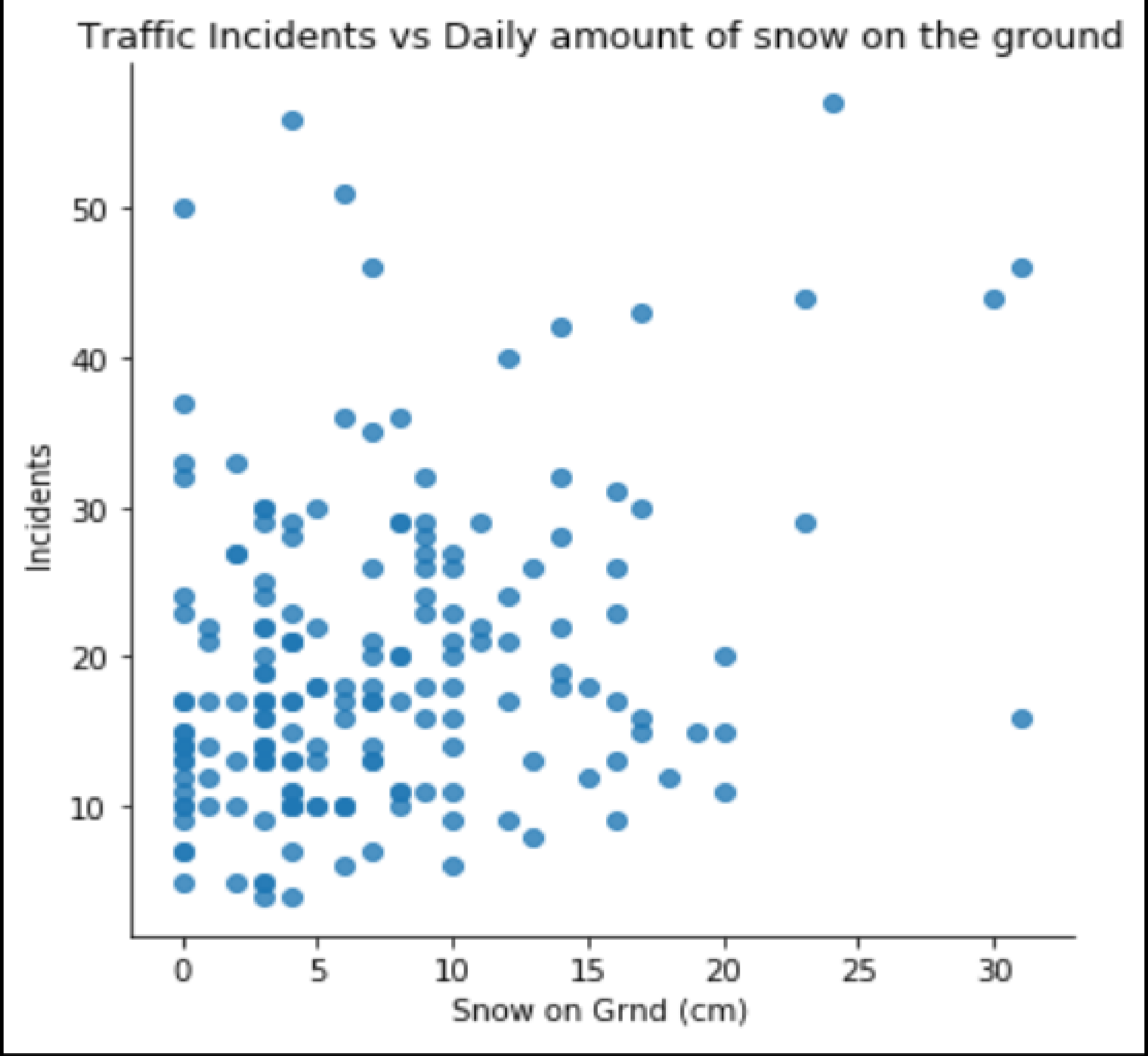
```
The Spearman correlation of incidents and mean temperature is:  0.069695986840367
```



The Spearman correlation of incidents and total daily precip is: -0.299331622688534

Hard Data Visualization – Plots

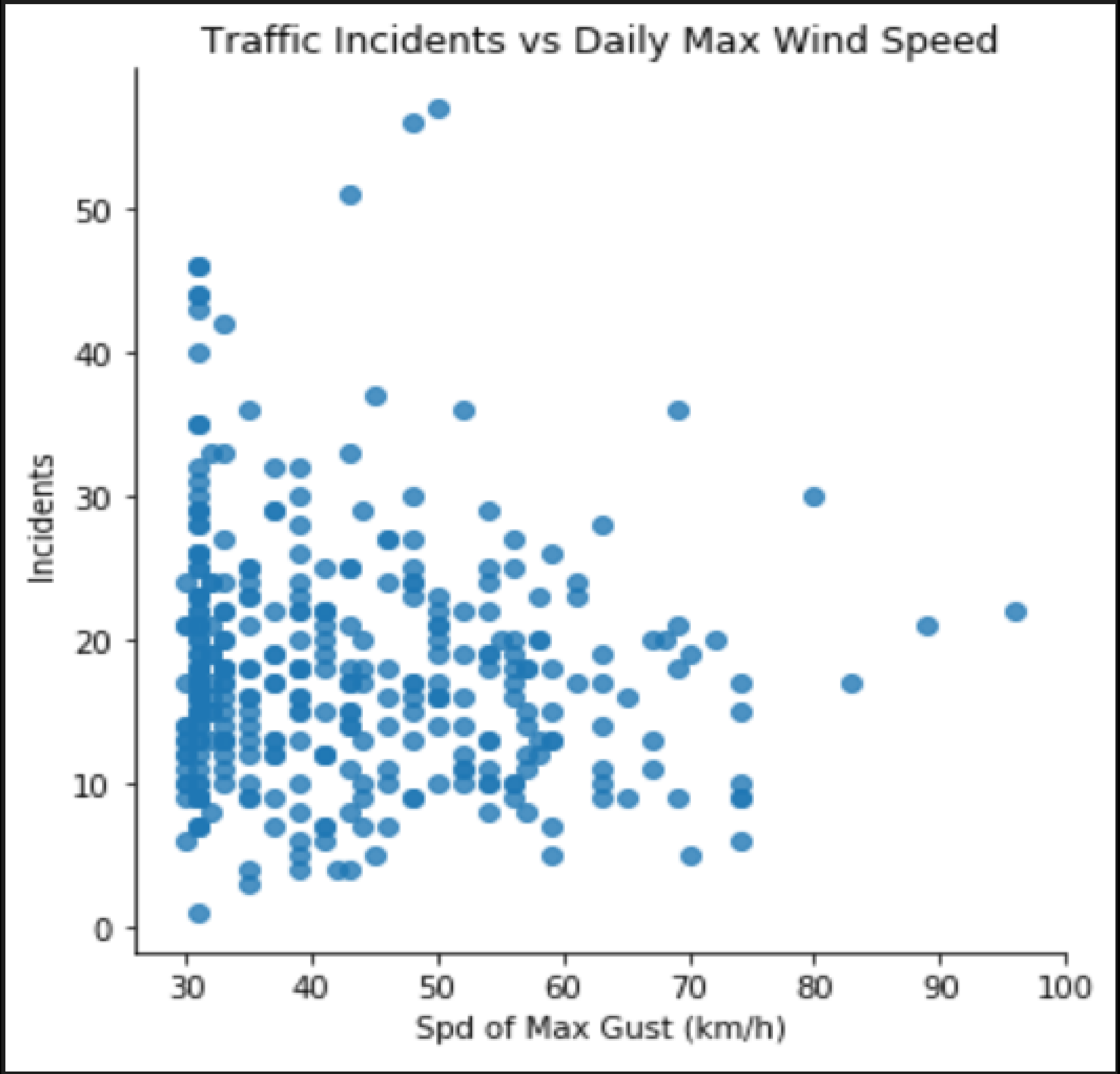
- There is weak correlation between the total daily precipitation and the number of traffic incidents.



Hard Data Visualization – Plots

- There is no significant correlation between the daily amount of snow on ground and the number of traffic incidents.

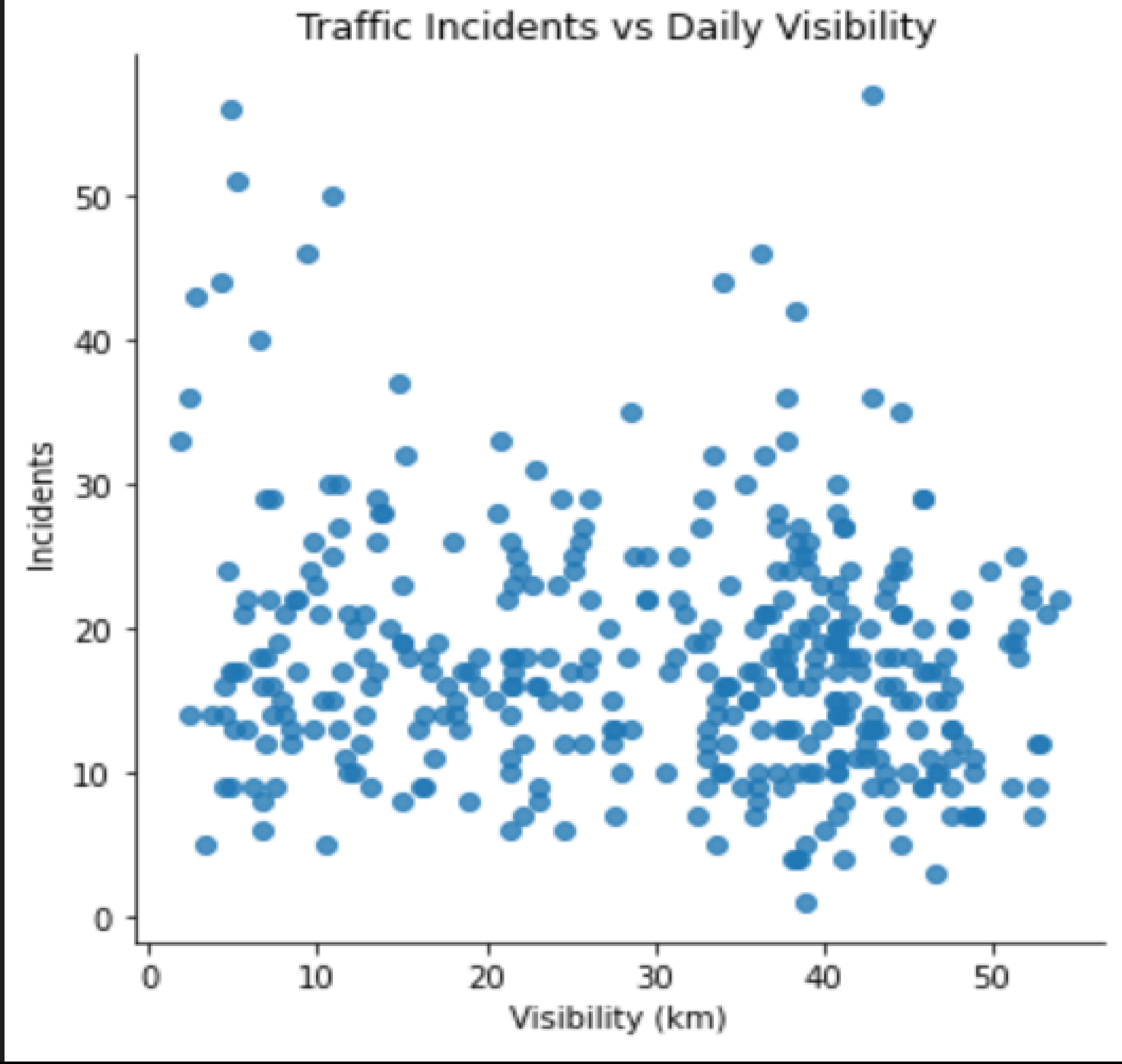
```
The Spearman correlation of incidents and snow on ground is:  -0.08631998990557002
```



Hard Data Visualization – Plots

- There is weak correlation between the daily max wind gust speed and the number of traffic incidents.

```
The Spearman correlation of incidents and max wind gust is:  0.14462157376758245
```



Hard Data Visualization – Plots

- There is no significant correlation between the daily visibility and the number of traffic incidents.

```
The Spearman correlation of incidents and daily visibility is:  -0.06074302726224958
```

Conclusion

Traffic speed limits were seen to have very weak correlation with incidents.

Weather features and traffic incidents were also found to have very weak correlation. Additionally, visibility was found to have very weak correlation with incidents. Their relationships can be seen in the graphs shown previously.

Traffic volume and incidents were found to have strong correlation, indicating that areas with more vehicles tend to have higher incident counts.

Personal insight: While the data and visualizations show strong correlation between traffic signals and signs, these are probably not causal in nature. A more realistic statement would be that areas with high incident counts tend to have more signs and signals because of this anomaly, not the other way around. Another data factor probably exists that would warrant additional analysis of this problem and would most likely be related to the infrastructure design of these problematic incident locations.