Quick Seattle Traffic Flow Analysis

Gavin Peterkin

October 2015

Abstract

This is a fun, short project that uses data available publicly from data.seattle.gov. I study changes in traffic flows on 270 arterial roadways located within and outside of the downtown area from the years 2007 to 2009. The primary purpose of this project is to figure out some basic information that's present in the data set while also becoming more familiar with R. I study the overall distribution of traffic counts, determine large trends in traffic flow, compare the downtown and out-of-town traffic flow, and perform a street level analysis.

1 Introduction

The data (available at the following URL: https://data.seattle.gov/Transportation/Traffic-Flow-Counts/7svg-ds5z) provide a character string for street name (STNAME), a boolean indicating whether the street is inside or outside of the downtown area (DOWNTOWN), a year, which varies from 2007 to 2009, and an integer average indicating the volume of traffic in both directions where the particular spot study was carried out. This value is referred to as the "seasonally adjusted average weekday daily total" (AAWDT), and, as the name suggests, it is an adjusted average that indicates the average number of vehicles that crossed the spot study on an average yearly 24 hour weekday.

By studying the changing yearly traffic flows on individual streets, one can determine which streets are more in need of upkeep or upgrade. One may also study changing traffic flows to determine which streets *will* be in need of upgrade. The data is also an invaluable asset in modeling large scale traffic flows on other independent variables. One can also study large patterns like traffic flow within the downtown area and outside of it.

It is also important to keep in mind that traffic flow is not equivalent to "traffic." It is possible to have a large flow of vehicles and absolutely no "traffic" if, for example, one considers an 8 lane highway. It can handle a large volume of vehicles without significant backups or delays. Unfortunately, the data set does not include the number of lanes on each street nor does it contain the particular location on the road of the spot study. These are glaring omissions!

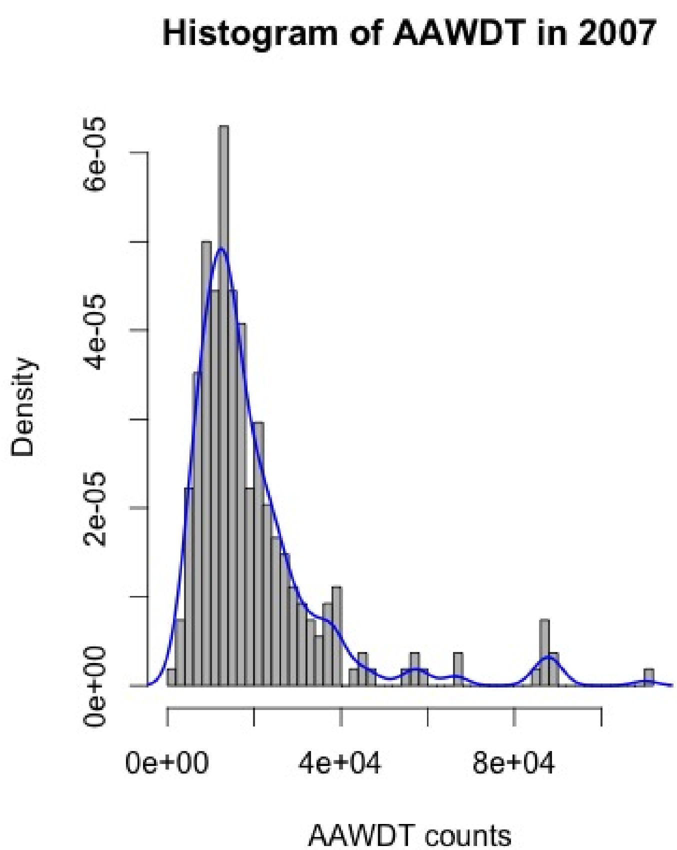
2 Methods & Analysis

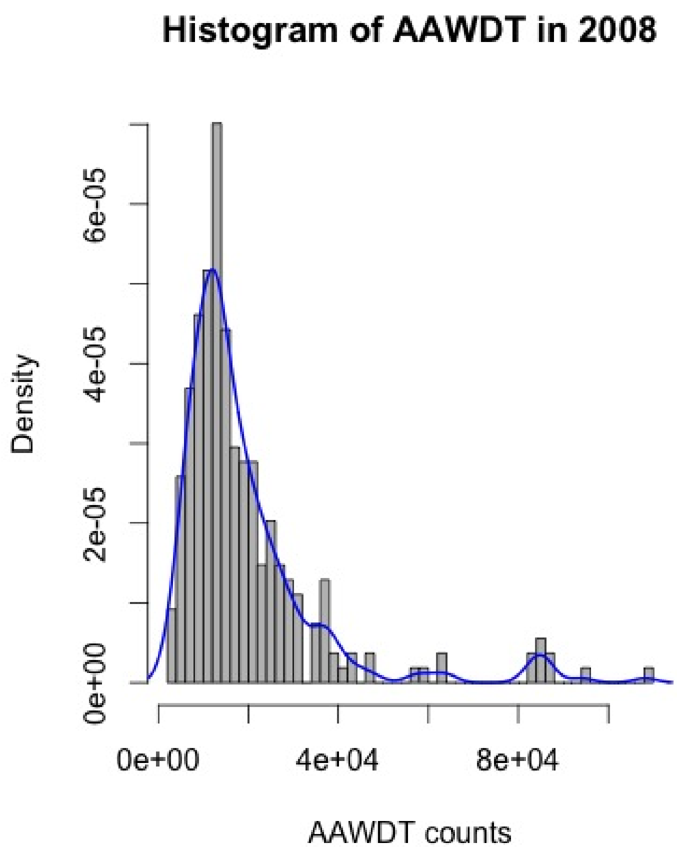
All data were analyzed and plotted in R using the following packages: DPLYR, GGVIS, GGPLOT2, and SN.

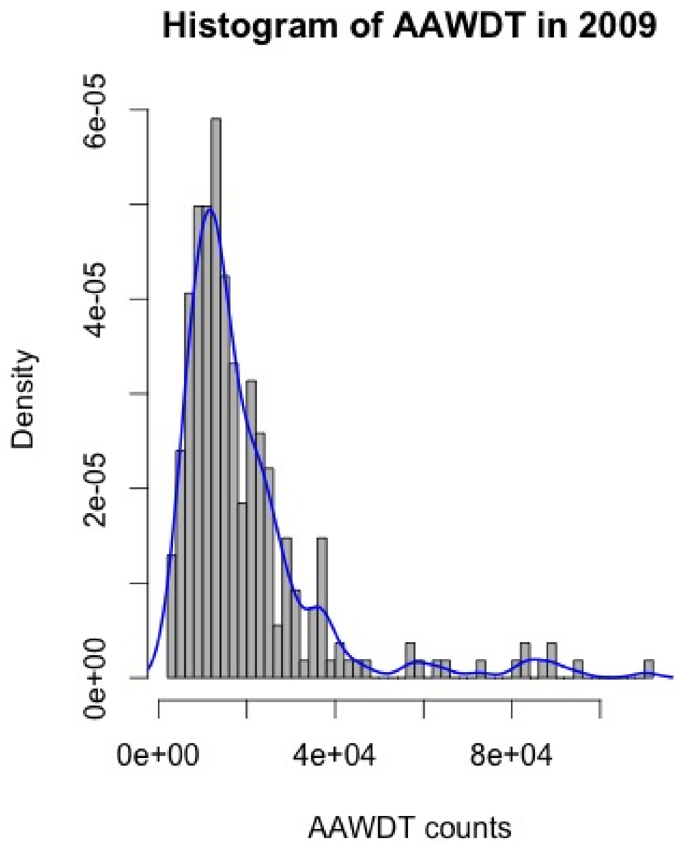
2.1 Initial Exploratory Analysis

The following histograms show how the overall distribution of AAWDT (traffic counts) changes over the three years of the study.

After an admittedly somewhat cursory sweep through different bin sizes, I settled on the size of 40 breaks because it provided a detailed overview of the distribution without over or under smoothing the data. What is somewhat interesting is the peak at around 90,000. The blue fitted line is based on the density() function, so it uses a gaussian kernel.




Histogram of AAWDT in 2008
Density
0e+00
2e-05
4e-05
6e-05
0e+00
4e+04
8e+04
AAWDT counts



One can see that most streets experience a flow of approximately 20,000 vehicles. It is still unclear how the distribution of traffic flow changed from 2007 to 2009, so it's worthwhile to determine some basic statistics.

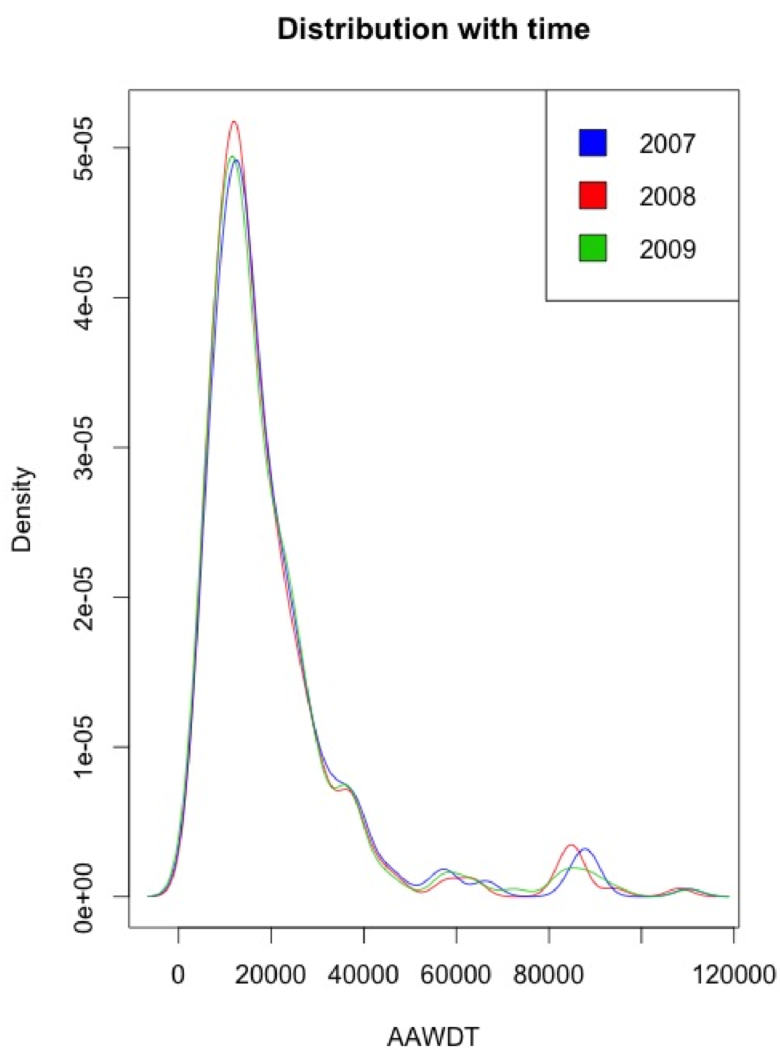
Year	Mean	Median	Std. Dev.	Max	Min	Total Count
2007	20,007	15,200	16,637	110,100	1,800	5,401,873
2008	19,532	14,700	16,529	108,200	2,100	5,293,300
2009	19,580	14,800	16,665	110,300	2,200	5,306,300

Δ Year	Δ Mean	Δ Median	Δ Std. Dev.	Δ Total Count
2008 - 2007	-475	-500	-108	-108,573
2009 - 2008	+48	+100	+136	13,000

Very generally speaking, for the average road there was less traffic flow in 2008 than there was in 2007. There was also less total traffic flow in 2008. In

2009, traffic flow once again increased, but it didn't seem to increase to 2007 levels. The decrease in traffic flow in 2008 is most likely due to the decreased economic activity as a result of the recession.

It's questionable how useful these very basic statistics are, so a more detailed look is necessary. The following figure plots once again the density functions over the three years.



Overall, this doesn't suggest anything remarkable. The results from before seem essentially correct. The use of a non-parametric gaussian kernel fit should also be questioned. (See the conclusion for more discussion on this point.)

There appears to be an interesting evolution of the peak at around AAWDT=90,000. It appears that its shape remains about the same from 2007 to 2008 although it's mean decreases. In 2009, its peak density decreases noticeably.

2.2 Downtown vs. surrounding area over time

The data set contains one chunk of information that can be very useful in figuring out if there's a correlation between traffic flow and overall economic activity. Over the three years studied, the total average traffic flow over all streets within the downtown area (DT) was always lower than the traffic flow outside of the downtown area (Not DT).

The following mean total traffic flows per street were averaged over streets DT or Not DT.

DT	2007	2008	2009
Mean total traffic flow/ST NOT DT	20,654	20,245	20,289
Mean total traffic flow/ST DT	15,659	14,726	14,800
Abs value of difference	4,995	5,520	5,489

It is not necessarily significant that the mean flow not DT is greater than the mean flow DT because the data set could be biased by the specific street selection (i.e. small downtown streets with limited flow and large out-of-town streets with a heavy flow). However, it is most likely the case that overall there is indeed a greater traffic flow outside of the downtown area. This is most likely due to the twice daily, weekday commute into and out of the city.

More significant than any of these concerns is how the traffic volume changed over this three year period. It seems as though DT traffic flow experienced greater changes in total AAWDT counts than the Not DT area. This change in traffic volume does seem to possibly correlate with the 2008-2009 recession because one may hypothesize that economic activity is more closely correlated with DT than it is with Not DT traffic.

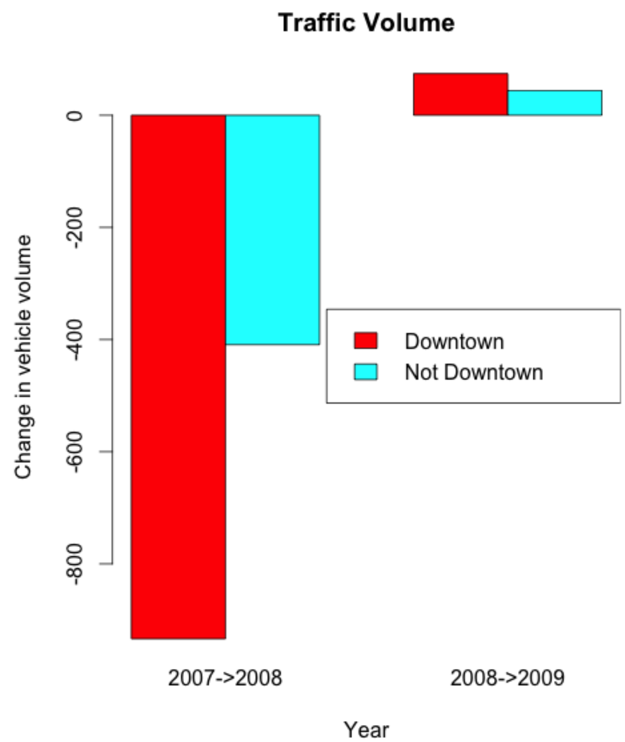


Figure 2.2 shows the changing traffic flow between the three years.

2.3 Street level analysis

Rather than merely determining which streets have the most or least flow, it may be worthwhile to determine which streets have the greatest changing flows. This analysis is far more meaningful because it naturally takes into account the fact that different streets have different capacities. For the following analysis, for each street I determine the change in AAWDT over both periods ('07 to '08 and '08 to '09) and calculate the percentage change in traffic volume.

From 2007 to 2008, the average percentage change in road use was -1.7%. In 2008 to 2009, the mean change was 2.3%. This roughly aligns with what we discovered earlier (a drop in 2008 and slight rise in 2009).

Streets with Increasing Flow I determine which streets have the greatest increasing traffic in the periods from 2007 to 2008 and from 2008 to 2009. These are roadways that have experienced an increase in use of at least 20%.

2007 to 2008

Street	Total Increase	Percentage Change
NICKERSON ST	11,900	100.8
SW 106TH ST	6,400	66.6
S HENDERSON ST	3,200	47.1
NE 45TH ST	8,800	25.6
ML KING JR WAY S	5,100	25.6
WEST SEATTLE BR EB & WB	15,800	23.8
MLKING JR BLVD S NB & SB	3,400	22.9
RENTON AVE S	1,500	20.8
EAST MARGINAL WAY S	2,700	20.4

2008 to 2009

Street	Total Increase	Percentage Change
BALLARD BRIDGE	51,300	869.5
MELROSE AVE ON RAMP	10,800	71.0
ALKI AVE SW	2,600	36.6
RENTON AVE S	2,200	24.1
ML KING JR WAY S	3,700	20.7

Streets with Decreasing Flow The following streets have flows that have decreased by more than 20%.

2007 to 2008

Street	Total Decrease	Percentage Change
BALLARD BR	51,800	-89.8
EAST MARGINAL WAY S	5,200	-29.9
RENTON AVE S	3,000	-24.8
15TH AVE NE	3,800	-23.3

2008 to 2009

Street	Total Decrease	Percentage Change
SW 106TH ST	7,200	-45.0
S JACKSON ST	2,800	-22.4
EAST MARGINAL WAY S	3,500	-22.0
14TH AVE S	3,700	-20.7

3 Conclusion

3.1 Summary of Results

The overall traffic flow seems like it may be roughly correlated with the level of overall economic activity, and DT traffic is more closely correlated. This is something I will continue to review in the future. The street level results suggest which roads may be in need of expansion or renovation due to a large increase in flow and which streets may be in need of further study because of large decreases in AAWDT count.

3.2 Ideas for the future

Instead of using the built-in density function, I think it would be worthwhile to instead study the histogram distribution with a a more suitable function chosen on the basis of a model (e.g. a parametric model). In other words, what would the distribution look like in an idealized world. Intuitively, I would guess a right skewed normal distribution, but the solution may actually be more complex due to the distribution of the number of lanes per road or other factors.

A natural limit occurs at $AAWDT = 0$ on the left sides of the histogram curves.

Luckily, there's an r package, "sn", by Adelchi Azzalini that fits skewed normal distributions. Assuming a skew normal model and a maximum likelihood estimation fitted to a probability density histogram of AAWDT, I find the following:

Year	Mean	Standard Deviation	γ
2007	1.36e-5	1.05e-5	0.195
2008	1.39e-5	1.07e-5	0.995
2009	1.37e-5	1.05e-5	0.995

This suggests an increase in the mean in 2008 and a decrease in 2009. To further consider the possible correlation with economic activity one may look at the GDP of the USA and Seattle Metro Area. I will continue working a skew normal distribution to model traffic flows in the future.

Year	GDP Seattle (Billion $)	Seattle Pop.	GDP per cap USA ($)
2007	235.44	592,647	48,061.42
2008	242.81	602,943	48,401.49
2009	240.14	616,627	47,001.43

*Sources:
http://www.statista.com/statistics/183863/gdp-of-the-seattle-metro-area/
World Bank, U.S. Census Bureau.