Report on Edinburgh
Bike Count Data

Group members:
Sally Bai,
Yao Tong,
Xuanchen Yao

Table of Content

Introduction

1. Project Context

The City of Edinburgh Council (CEC) has brought value to improve public transportation in order to achieve a cycling and walking-friendly city, and more importantly, to enhance people's live quality. To approach citizens' cycling behaviour, path usage and traffic density, the Bike counters were installed to record hour-based bike counts.

In addition to the data collection of CEC cycle team, the organization "Spookes" has also installed counters for promoting cycling related projects. Our work will contribute to previous existing datasets from CEC cycle team to give the hypothesis and compelling statistics.

2. Aims and Achievements

The main purpose for this project is to approach people's cycling habits in Edinburgh city - how did people perform in different time periods and conditions of weather or places. Divided into different time session along the year 2015 and 2016 of full of visualization, the datasets which provided by the city council of Edinburgh have shown interesting findings.

There are a series of questions we aim to answer when dealing with data:

Are there any irregular data worthy of exploring? What would be the potential cause?

How did different time periods effect on cycling traffic? Seasonality? Weekdays and weekends?

Are there any relationships between bike counts and variables such as weather and altitude?

Data Preparation and Processing

1. Data Description

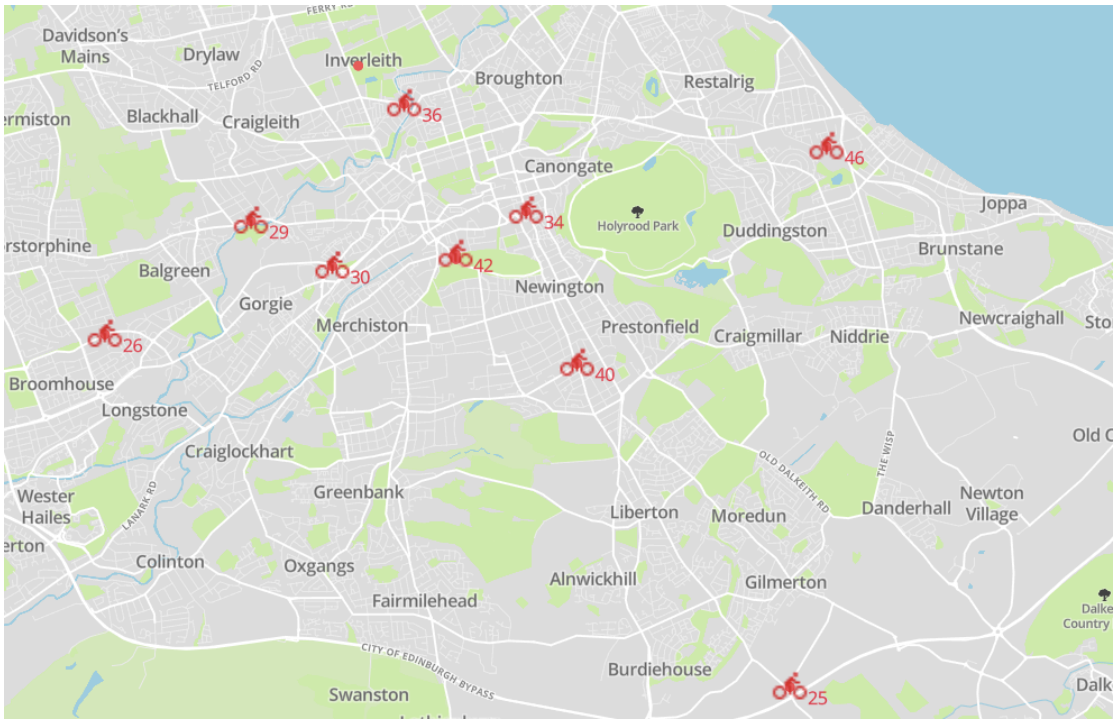
The data are from 19 off-road counters and 29 on-road counters from 2004 to 2016. Most of the data has been collected constantly, only the data temporarily missing are due to battery failure which also shows an irregularity in graphs. Our project chose 13 sites out of 48 owning to they share a common data collection time period between 22/03/2015 and 18/06/2016.

There are total 5 columns, representing

counter_id: the data is collected via labelled automatic counter (1-48);
date: the day on which the data was collected (18/03/2010);
time: the hour on which the data was collected (0~23);
channel_1: the direction of travel (north-bound);
channel_2: the direction of travel (south-bound)

2, Data Preparation

Each bike count site was collected in weeks. First step is to combine the site data according to their count id. Then the datasets were renamed according to their locations.



https://goo.gl/ExjNPN

*The figure indicates the distribution of 13 bike counters according to their coordinator

Data Analysis Methodology

All the data analysis and visualisation have implemented through ipython notebook. Additional libraries such as numpy and pandas have also been used. The format of datasets has also been altered to cater different time periods.

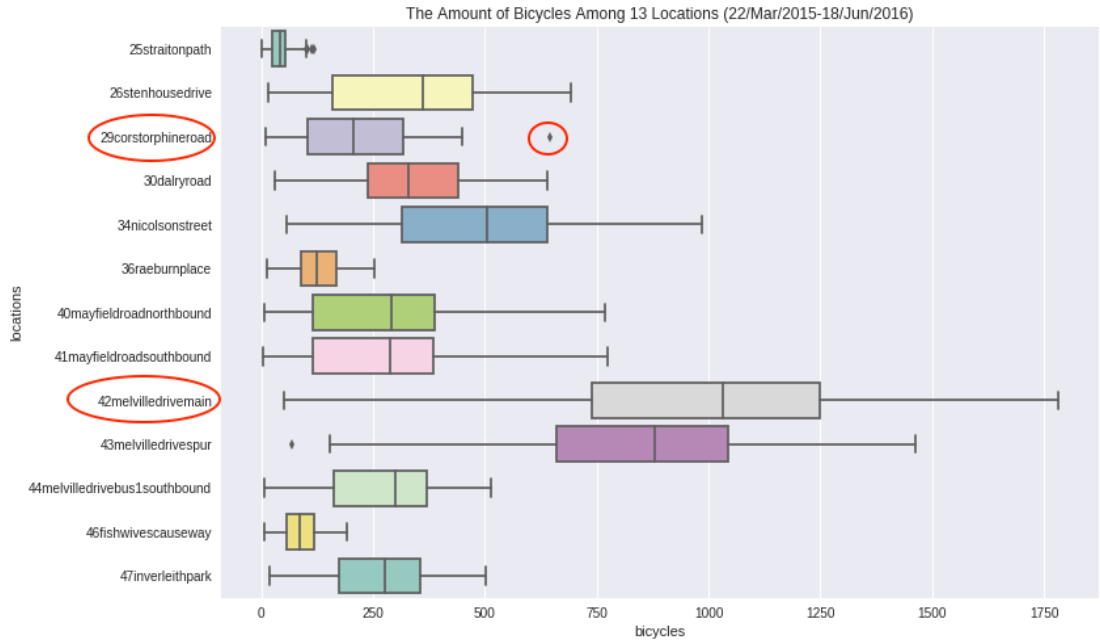
1. The traffic of channels is first added up. The change of bike counts can be presented in distribution plot, boxplot, regression plot and so on.
2. Parameter of time period
   the total traffic of each sites is added up to visualise the change within a day.
3. Other datasets: relate weather data to cycling data with time, and probe the linear relationship.
   https://www.ed.ac.uk/geosciences/weather-station/weather-station-data
4. Other datasets: approach the location with its traffic.
   https://www.google.co.uk/maps/@55.9445594,-3.1984787,14z?hl=zh-CN
5. Other datasets: combine altitude and bicycles data together to explore the relationship between them.

http://elevationmap.net/26-easter-belmont-rd-edinburgh-eh12-6ex-uk?latlngs=(55.94670067561135%2C-3.2567596435546875)

Results on Detected trends

Plots shown below present various forms of clustering on bike counts. Time line is one of the axis to parse bike datasets.

1) Analysis on cycling traffic at different locations



Comments:

- The spread of traffic distribution has shown different statistic range.

The interquartile of Site 42 and 43 are greater than other sites.

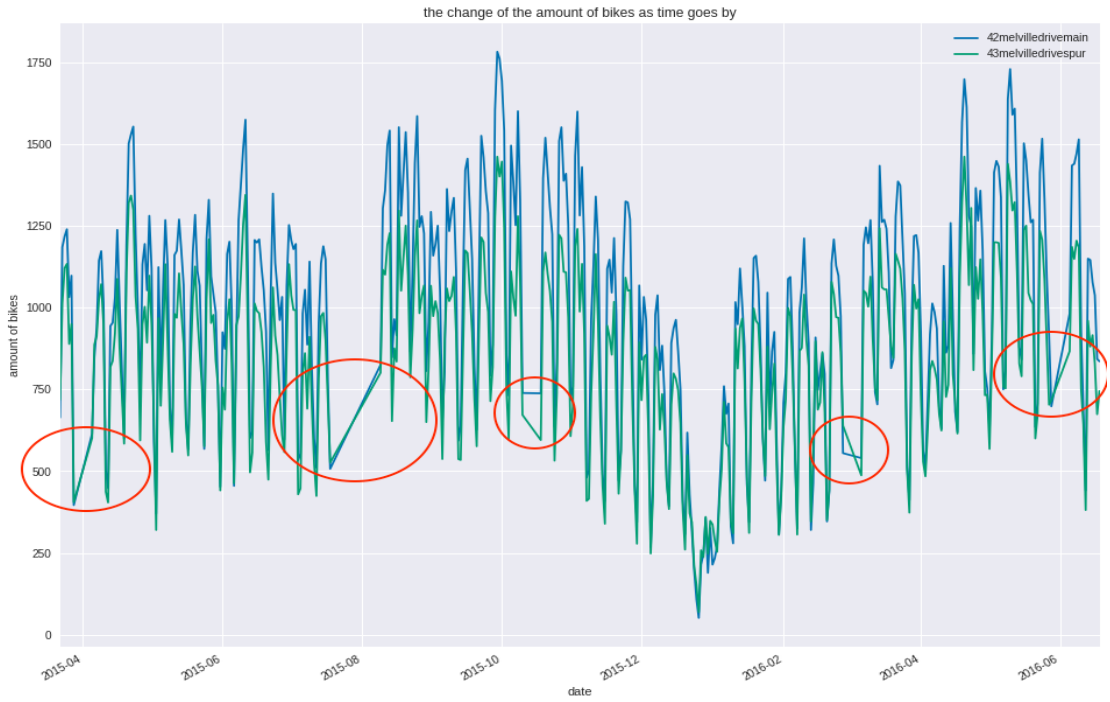
- The strange outlier appears in counter 29

After searching online with the key words like '2015 Sep 6th' and 'bike', we found that there was an event happening right there then, which is just crossing the counter set at Corstorphine Road, namely, 29corstorphineroad.

Pedal for Scotland- Scotland's Biggest Bike Event – 5th, 6th, 13th September 2015. We can find the following link for further information about what happened during that time:
http://www.dailyrecord.co.uk/lifestyle/health-fitness/pedal-scotland-scotlands-biggest-bike-6103718

According to these findings, bike events can also affect the amount of bikes significantly.

2) The change of traffic from 04/2015 to 06/2016 (take counter 42/43 as example)



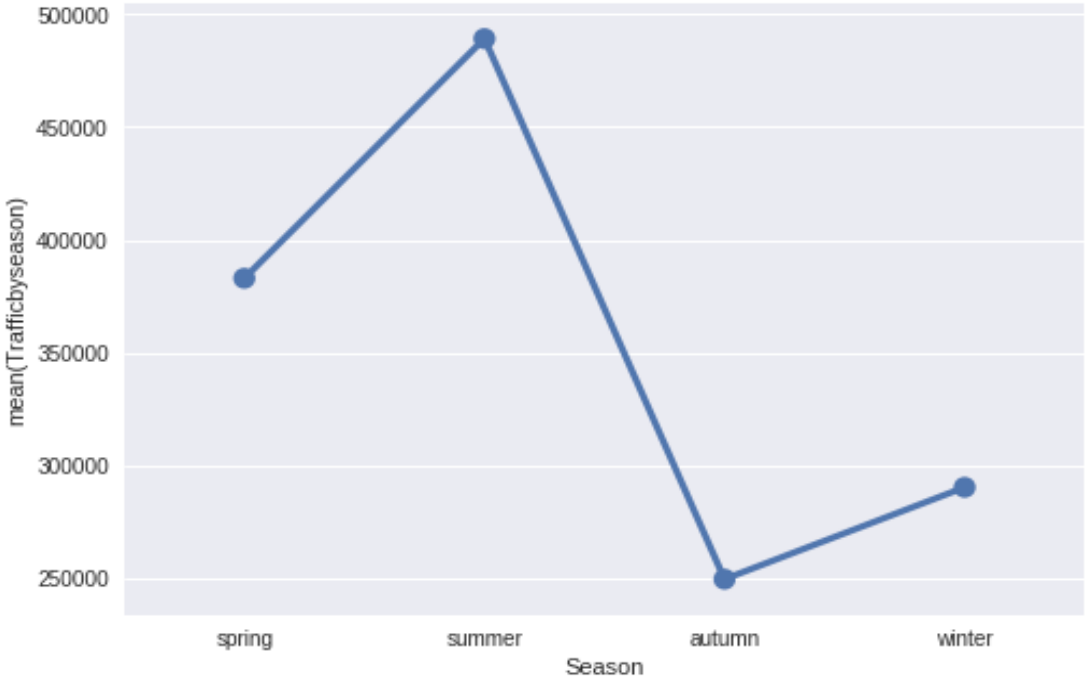
comments:

- Unlike other fluctuations, the steady and straight increases appear in the red circles. The bike counters are considered out of service;

- The bike counts over the year has fluctuated dramatically, and the peaked ups and downs indicated the different performance in days. In 2016/01 or so, the traffic hit its lowest point. To get a clearer point of view, we narrow down the time line:

  data reached its bottom around Dec.26, which fit in line with the Christmas.

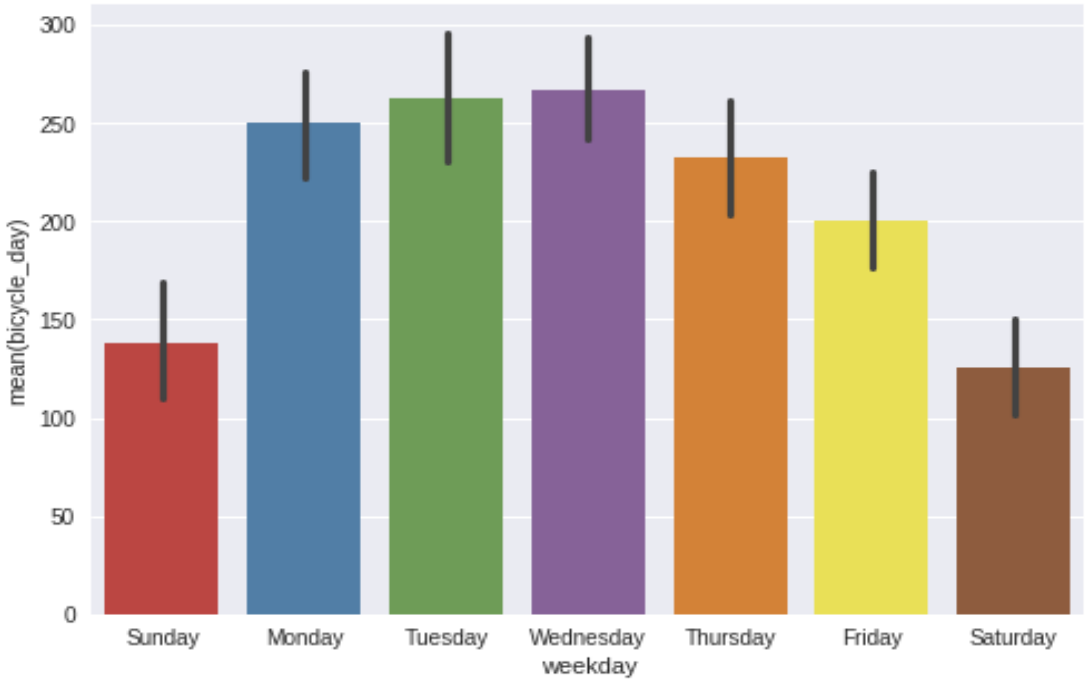
3) The general overview of seasonality cycling traffic



Comments:

- The bike counts peak in summer which are in accordance with common knowledge. To look into the specific determiner in seasonal factors, the relationship between weather and bike counts is analysed as followed.

4) Approaching the performance of bike counts in days of a week

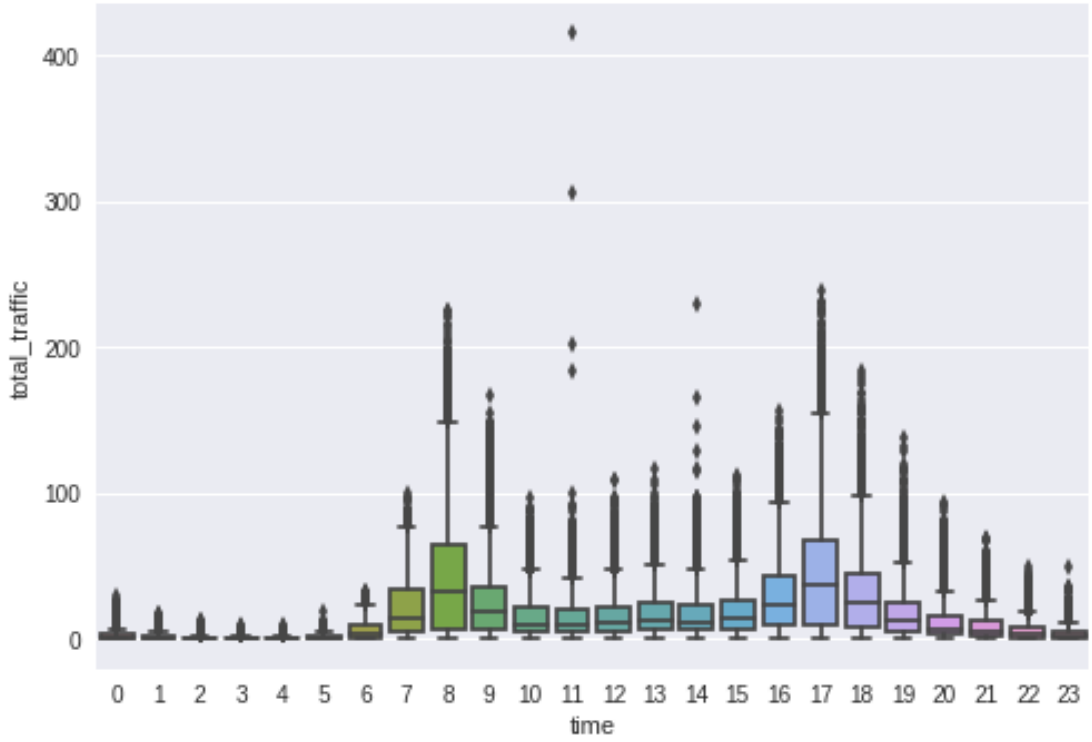


comments:

- People prefer to go out for biking on weekday rather than weekends, even taking confidence intervals into consideration

  We get the hint of the decrease in weekends, which the determined factor of bike traffic pertains to commuting. To peek into the traffic change within one day, the boxplot below shows the dispersion of traffic in an hour basis:

5) The performance of cycling traffic in a day



Comments:

- According to the graph, two rises occur in 7 to 9 and 16 to 18. It proves that the main traffic in weekdays is for commuting.
- The graph shows a significant cycling behaviour which cycling serves as the daily transportation. To advocate more cycling, more attention should focus on the cycling route to make commuting more convenient.

6) Other exploratory variables

- The correlations between weather and bicycle

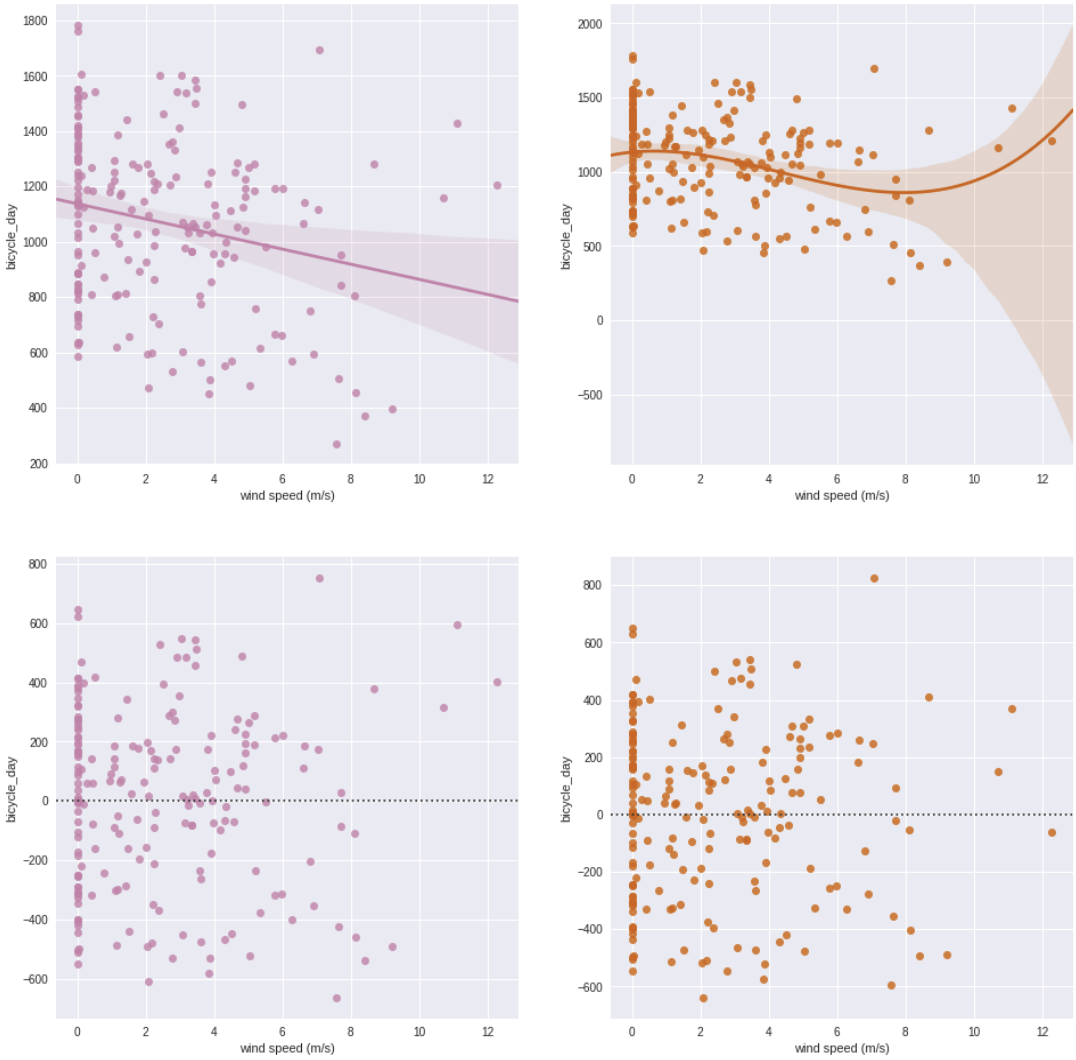
The plot below shows a weak linear correlation.

In accordance with it, the statistics is

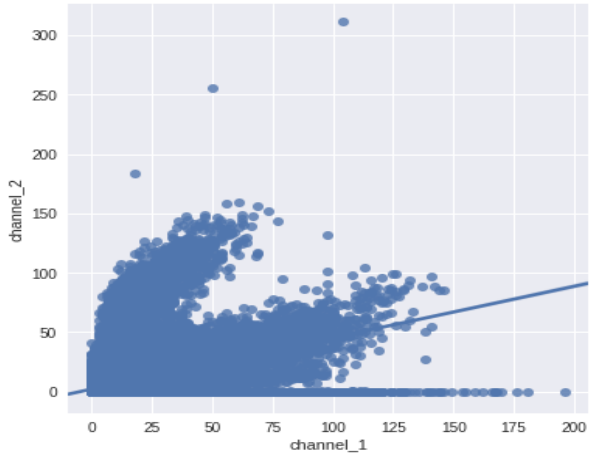
i. rvaule = -0.23 -- Wind speed does have a weak negative relationship with the bike counts as rvalue is between -0.3 and 0.

ii. pvalue = 0.001 -- We can accept our hypothesise as pvaule is less than 0.05.

iii. intercept = 1137 -- When wind speed is 0, bike amount is likely to be 1137.



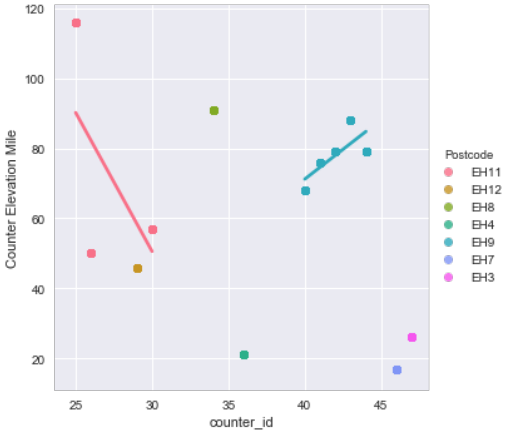
- The relationship between two route directions



Comments:
Plot doesn't show a strong linear relationship. It indicates that people might choose different routes in a day.

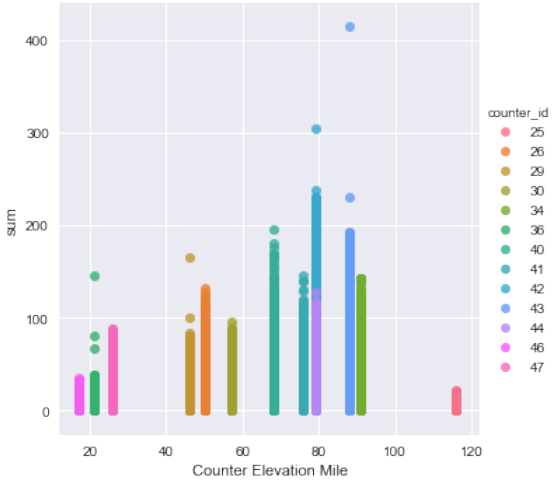
- The relationship between altitude and bike traffic data

1) The counters and its elevation and postcode



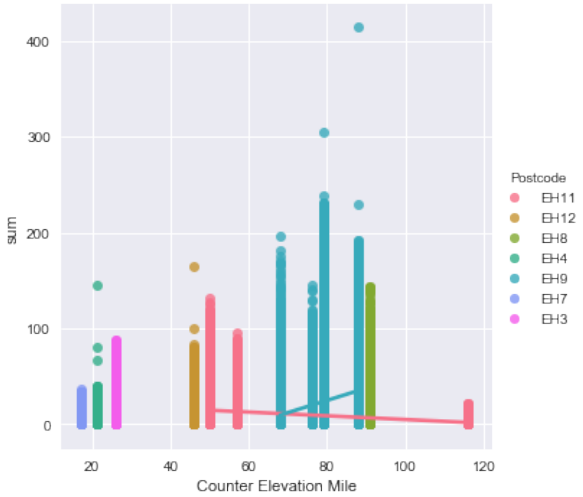
- The plot below shows investigations of data with counters and its elevation and the correlation of its location and bike flow of two channels.

2) Elevation and counter data sum (labelled with counter id)



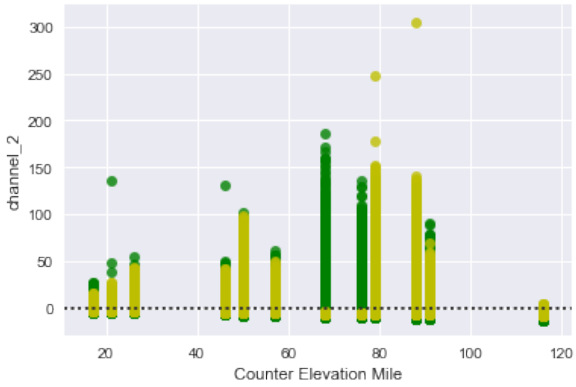
Comments:
We can learn from the graph at the left, the traffic flow is generally busy on the elevation of 70 to 90 in Edinburgh, which also have been cataloged in postcode as followed.

3) Elevation and counter data sum (labelled with its postcode)



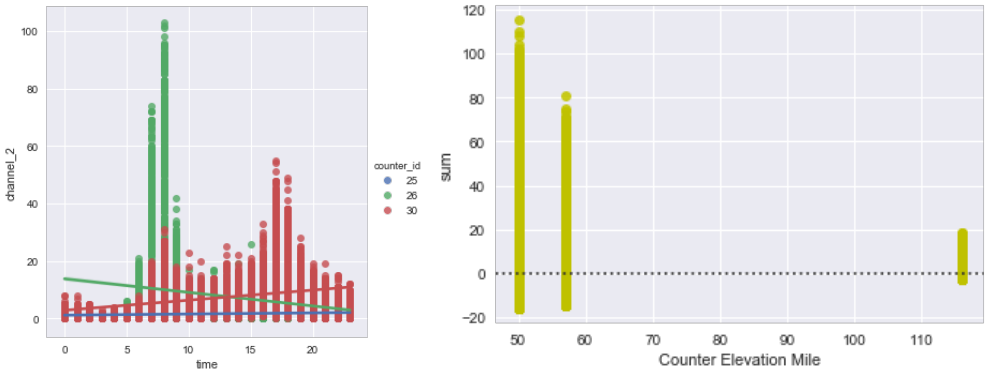
Comments:
The district of EH9, EH8 and Eh11 all display a high amount of traffic data and EH9, EH8 are in the center of city, EH11 locate in west of city. In next part I explored the relation between its channel(direction) and its time in these districts.

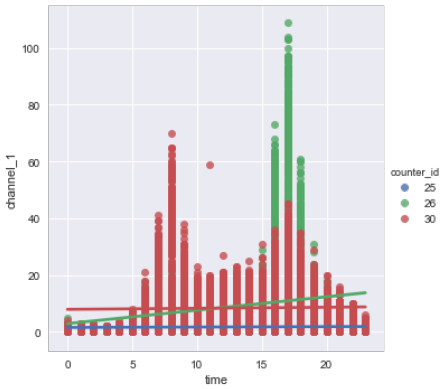
4) Elevation and 2 channels' data



Comments:
Although I found at the elevation around 70 to 75 have huge gap between two channels, its still lack of other evidence to make a statement of cyclists' biking behavior.

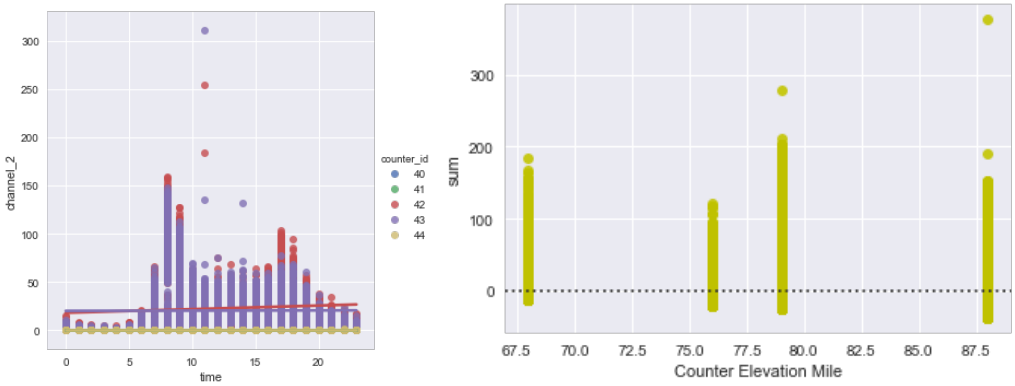
5) Elevation and 2 channels' data in EH11 (West of city)

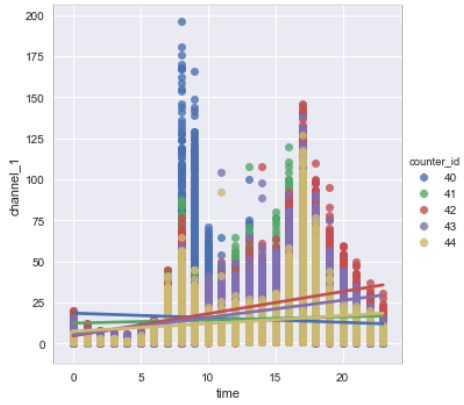




Comments:
Here we can find a significant comparison of No.25 and No.30 in its different directions in west of Edinburgh, in morning shown large amount of people going north at No.30 and afternoon going south. For No.25 counter, it is the opposite situation.

5) Elevation and 2 channels' data in EH9 (Centre of city)





Comments:
Although still exist many distractions of missing data of No.40 and No.44 counters in channel 2 (south-bound), we can still recognize the peak of traffic flow in morning and afternoon. The No.42 shows busier direction to south in morning and the other way round in afternoon.

Conclusion

Our purpose for this project consisted in trying to approach the cycling status in Edinburgh by means of dataset analysis. We adopted the “Extract”, “Transform”, “Load” process to parse the cycling traffic over a number of time periods. Bike counts in weekdays are greater than in weekends, so people choose taking bike as commuting transportation. The specific factor such as bike campaign and Christmas day will rise and descend the bike traffic respectively.

Moreover, we focused on investigating the underlying pattern by combining other open datasets such as the weather dataset of city Edinburgh and altitude distribution. While the correlation coefficient of those variables is not as strong as chronic analysis, it revealed the potential features which stimulated more in-depth examination.

Appendix

1. GitHub https://github.com/YaoTong9190/Bicycling
2. Website http://www.dudedu.com/web/index.html
3. Slides https://github.com/YaoTong9190/Bicycling/tree/master/presentation