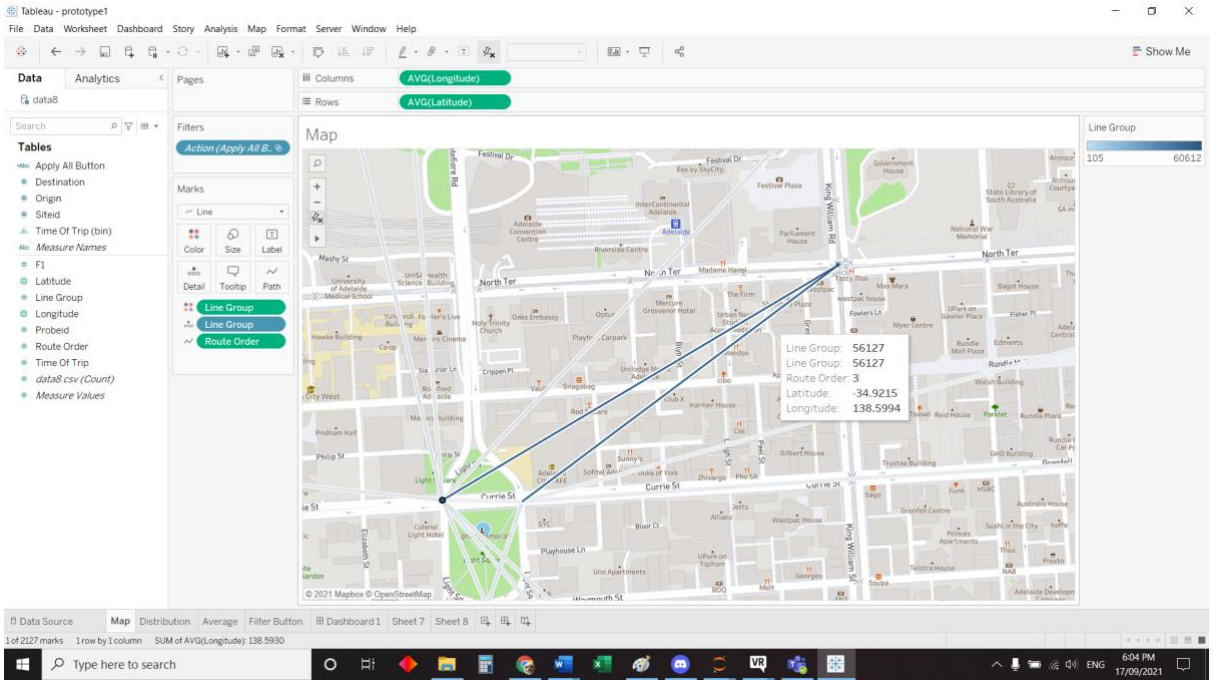
Constructing full trip car routes

1. For probe ids that are only seen once in the data are removed.
    - This is because we can not form any trip information from it.
2. Group the data by probeid and then sort by firstseenat.
    - This will show all the trip/s that cars have made
3. Then for each consecutive siteid that are previously grouped, find the time difference between them. Preferably in seconds so that things are easier.
    - This creates distributions of travel time for each consecutive siteid possible
4. For travel times that have been found to be outside -3 or +3 std they are marked as not a trip
    - This allows us to identify weakly when the trip information for one car ends and another starts.
        i. Later on, we explicitly mark the end between end and start of trips between different cars
    - As well unusual behaviors that are from cars staying in one place too long or going through it too quick which as an optimist, we could expect it to not be often if it is an outlier.
5. Based on trips that are marked as not being a trip based on the previous step, we fill in the gaps between to indicate if it is the start, journey or end of a full trip.
6. Once we have found the full trips we once again assign grouping to each full trip based on their origin and destination and reperform the analysis to mark full trips as outliers or not based on travel times between -3 and +3 std.
    - Outlier full trips can simply be removed assuming in the future we would have enough normal trips anyway
    - Full trips that are round trips may also be removed for the same reason
    - We could however in the future work on dividing up these trips based on arbitrary rules such as per 5min or per however many siteids but the code will be more complicated.

There is one thing to note, it is that sometimes when a car passes through a site it is not recorded which graphically looks like a straight line instead of a more pathed route as in the image below:

- Line group 56127 represents one full trip from siteid 3031 to 3030 which only passes through 3 unique siteids (data8).
- Looking at the original data where we had only grouped the cars and ordered by time the same is seen with car only detect across 3 unique siteids (data1)
- We could assume that sometimes the siteids do not detect a car passing by as given the map below the car should have gone through not just 3 unique siteids.



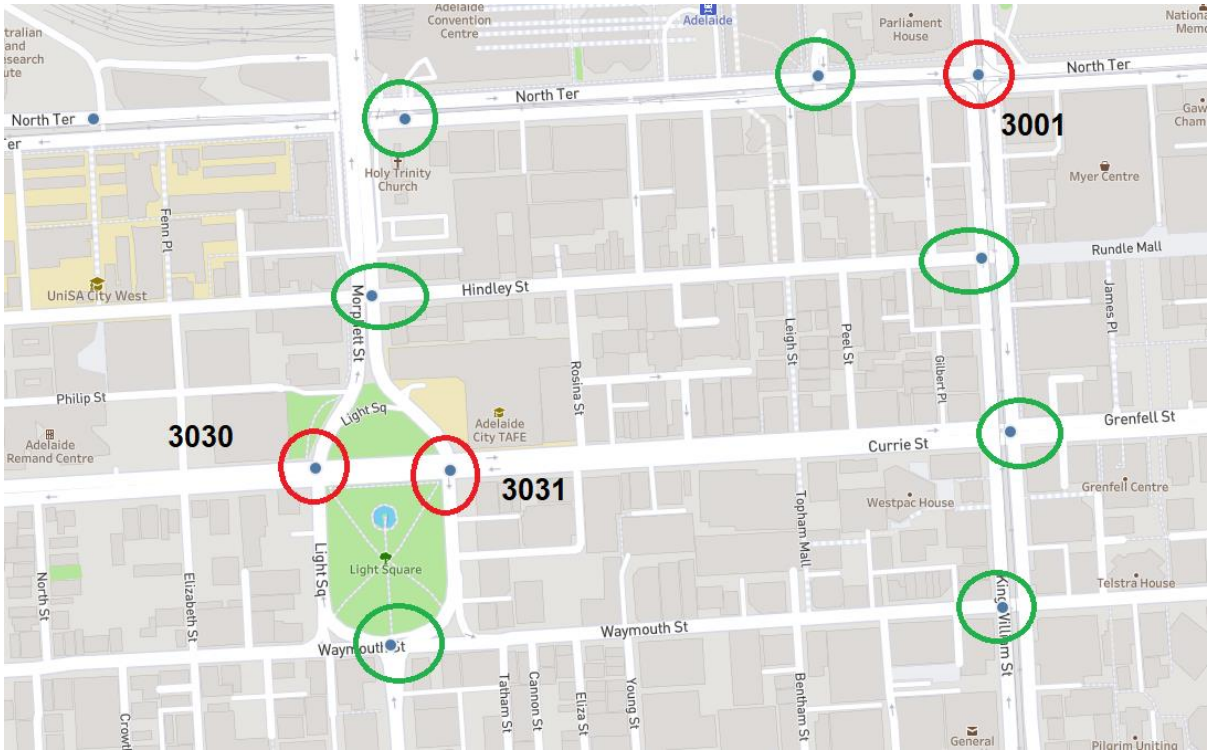

```
test = pd.read_csv('data8.csv')
```

```
test[test.line_group==56127]
```

	Unnamed: 0	probeid	siteid	line_group	route_order	longitude	latitude	Origin	Destination	time_of_trip
462815	462815	308393204	3031	56127	1	138.594261	-34.924693	3031	3030	2053
462816	462816	308393204	3031	56127	2	138.594261	-34.924693	3031	3030	2053
462817	462817	308393204	3001	56127	3	138.599399	-34.921541	3031	3030	2053
462818	462818	308393204	3030	56127	4	138.592959	-34.924676	3031	3030	2053

```
data1[data1.probeid==308393204]
```

	probeid	siteid	firstseenat	time_in_range	longitude	latitude
2320121	308393204	3031	2021-09-01 20:18:30+09:30	0 days	138.594261	-34.924693
2342823	308393204	3031	2021-09-01 20:19:03+09:30	0 days	138.594261	-34.924693
1859255	308393204	3001	2021-09-01 20:25:42+09:30	0 days	138.599399	-34.921541
5339108	308393204	3030	2021-09-01 20:51:53+09:30	0 days	138.592959	-34.924676
5391848	308393204	3030	2021-09-01 20:52:43+09:30	0 days	138.592959	-34.924676


North Ter
North Ter
North Ter
3001
Holy Trinity Church
Myer Centre
Rundle Mall
Hindley St
UniSA City West
Morphett St
Fenn Pl
Philip St
Light Sq
3030
Adelaide Remand Centre
Adelaide City TAFE
Currie St
Grenfell St
3031
Westpac House
Grenfell Centre
Light Square
Light Sq
Waymouth St
Telstra House
King William St
Topham Mall
Rosina St
Leigh St
Peel St
Gilbert Pl
James Pl
North St
Elizabeth St
Tatham St
Cannon St
Eliza St
Young St
Bentham St
Parliament House
Adelaide Convention Centre
Adelaide