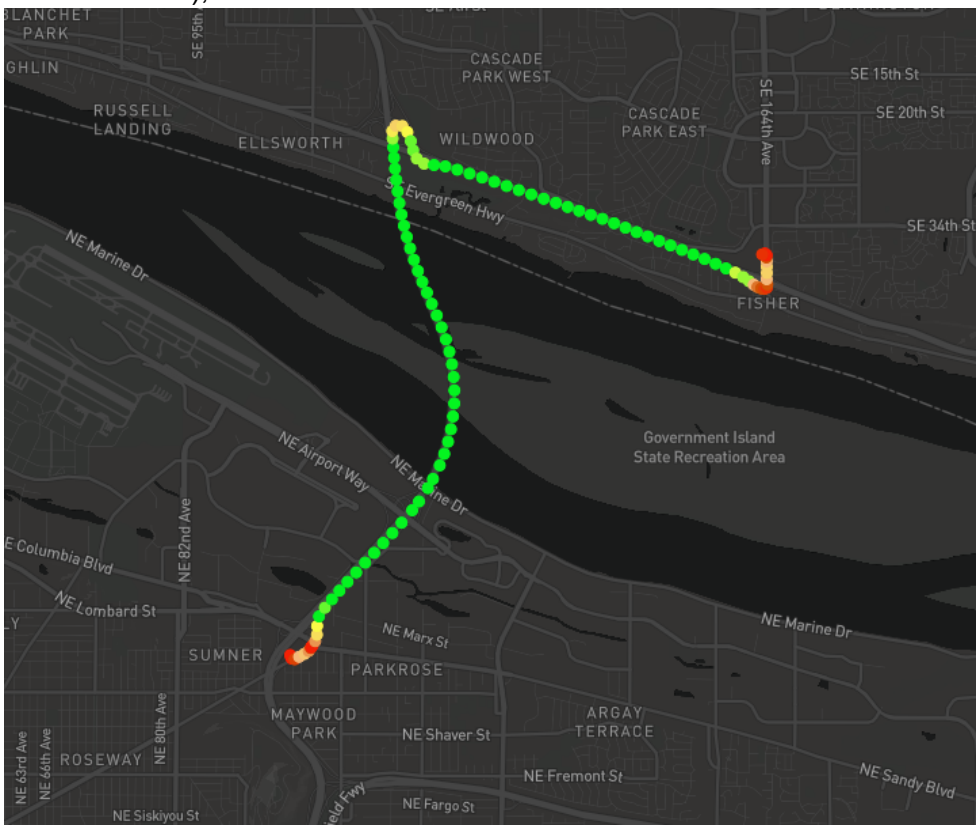
Michael Robertson
Data Engineering
CS410
Bruce Irvin
Genevieve Lalonde

Project Part 3 Integration

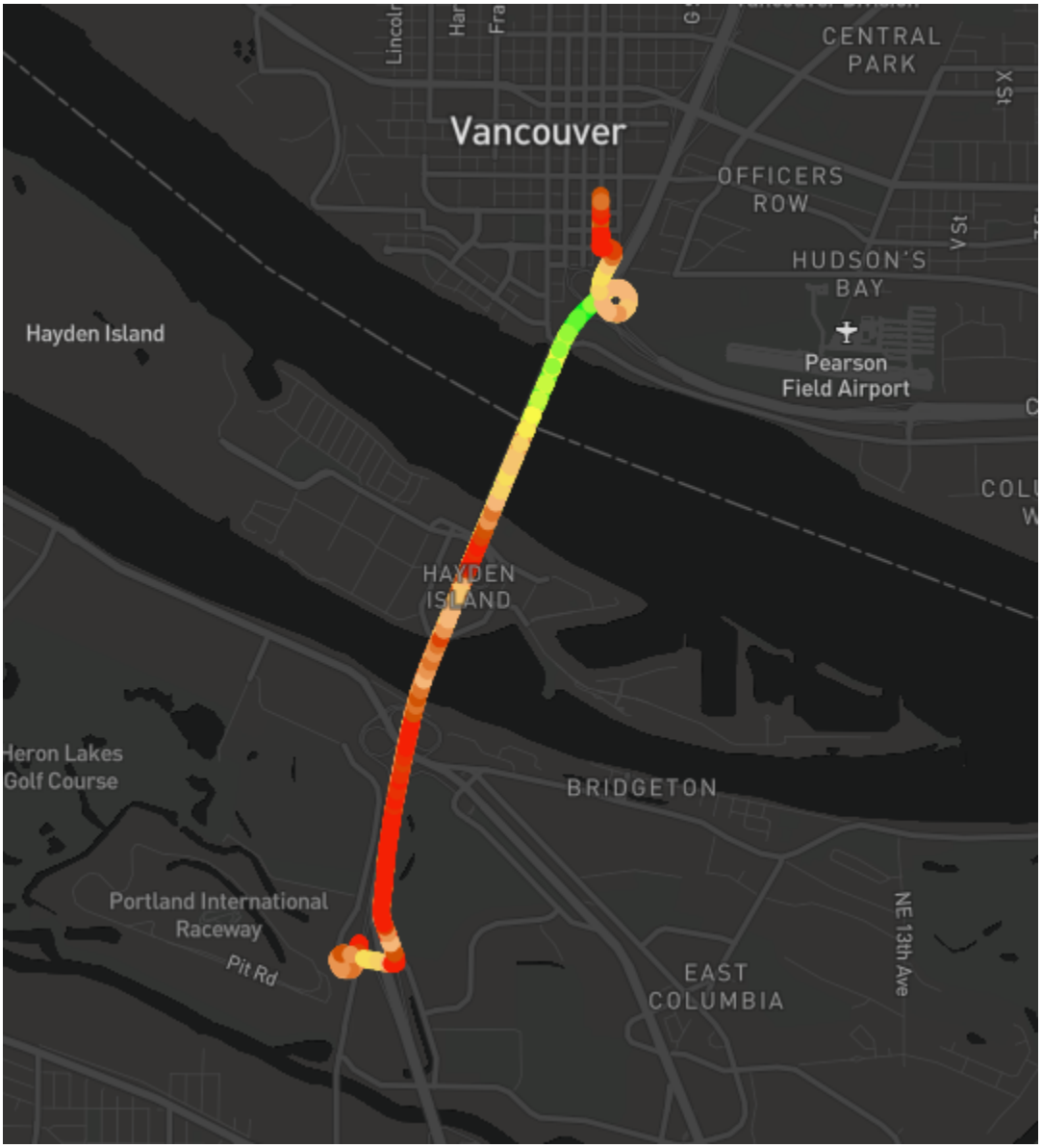
Visualization 1. A visualization of speeds for a single trip for any bus route that crosses the Glenn Jackson I-205 bridge. You choose the day, time and route for your selected trip. To find a trip that traverses this bridge, consider finding a trip that includes breadcrumb sensor points within this bounding box: [45.592404, -122.550711, 45.586158, -122.541270]. Any bus trip that includes breadcrumb points within that box either crosses the bridge or goes swimming in the Columbia river!

```
SELECT longitude, latitude, speed FROM breadcrumb JOIN (SELECT MAX(trip_id) FROM
breadcrumb WHERE DATE(tstamp)='2020-10-24' AND (longitude >= -122.550711 AND
longitude <= -122.541270) AND (latitude >= 45.586158 AND latitude <= 45.592404)) AS results
ON (trip_id = results.max);
```



Bus speeds for the breadcrumb data pertaining to the largest trip_id that was within the region of (45.592404, -122.550711) - (45.586158, -122.541270) on 24OCT2020.

Visualization 2. All outbound trips that occurred on route 65 on any Friday (you choose which Friday) between the hours of 4pm and 6pm.
SELECT B.longitude, B.latitude, B.speed FROM breadcrumb B JOIN (SELECT * FROM Trip WHERE service_key = 'Weekday' AND route_id = 60 AND direction = 'Out') as results ON (B.trip_id = results.trip_id) WHERE EXTRACT(hour from tstamp) >= 16 and EXTRACT(hour from tstamp) <= 18 AND DATE(tstamp) = '2020-10-23';



Bus speeds for the breadcrumb data of all outbound trips that are listed from route 65 on 23OCT2020 from 1600-1800.

Visualization 3. All outbound trips for route 65 on any Sunday morning (you choose which Sunday) between 9am and 11am.
SELECT B.longitude, B.latitude, B.speed FROM breadcrumb B JOIN (SELECT * FROM Trip WHERE service_key = 'Saturday' AND route_id = 60 AND direction = 'Out') as results ON

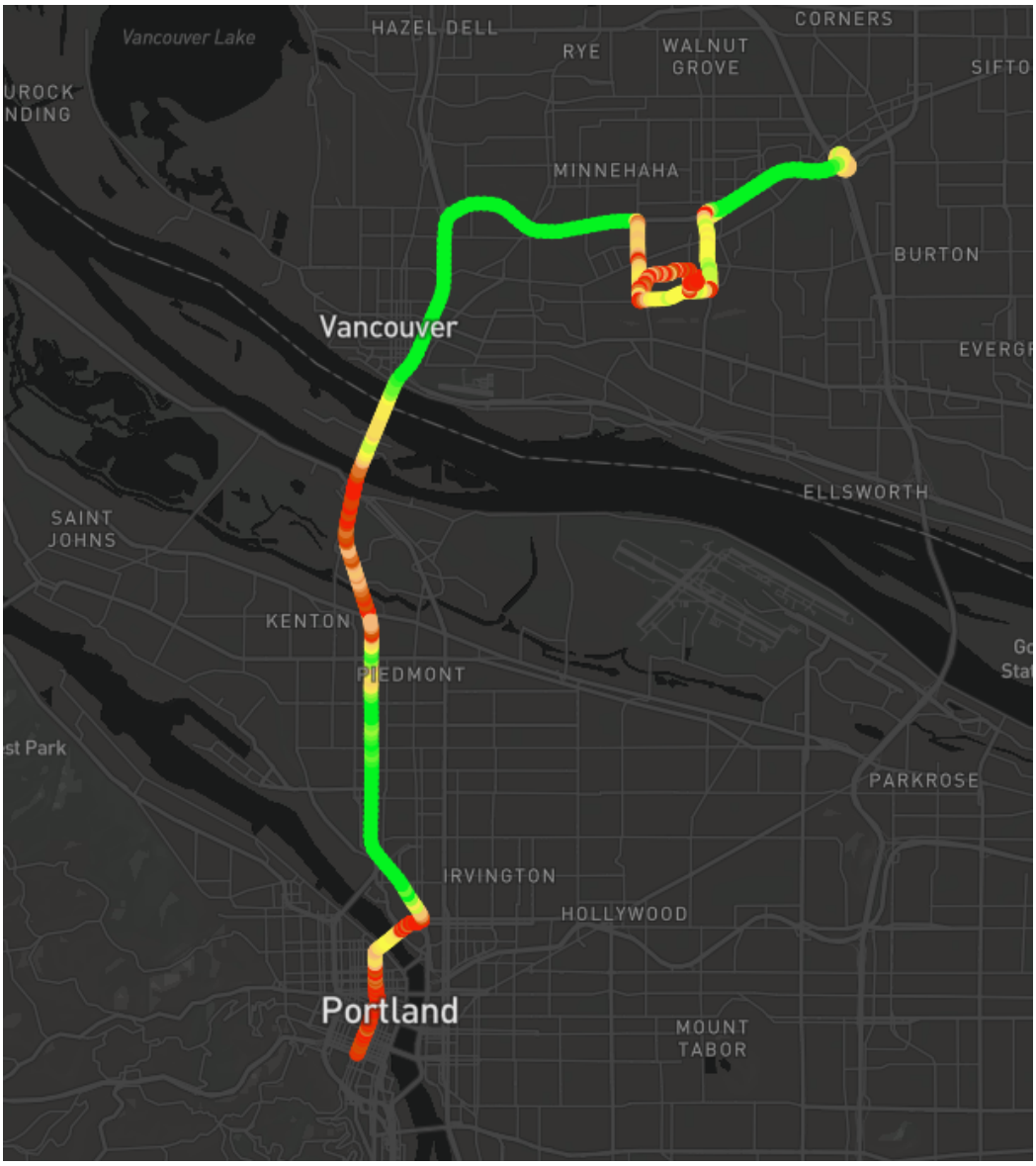
(B.trip_id = results.trip_id) WHERE EXTRACT(hour from tstamp) >= 9 and EXTRACT(hour from tstamp) <= 11 AND DATE(tstamp) = '2020-10-25';

NOTE: REALISED THAT WHEN CALCULATING WHAT SERVICE_KEY EACH DAY IS, ACCIDENTALLY MESSED UP THE INFORMATION THEREFORE THIS QUERY SAYS SATURDAY BUT THE DATE IS ACTUALLY FOR SUNDAY

Bus speeds for the breadcrumb data of all outbound trips that are listed from route 65 on 25OCT2020 from 0900-1100.

Visualization 4. The longest (as measured by time) trip in your entire data set. Indicate the date, route #, and trip ID of the trip along with a visualization showing the entire trip.

SELECT longitude, latitude, speed FROM breadcrumb JOIN (SELECT trip_id, (MAX(tstamp)::time - MIN(tstamp)::time) AS time, date(tstamp) FROM breadcrumb GROUP BY trip_id, date(tstamp) HAVING (MAX(tstamp)::time - MIN(tstamp)::time) = (SELECT MAX(results.time) FROM(SELECT (MAX(tstamp)::time - MIN(tstamp)::time) AS time, trip_id, DATE(tstamp) FROM breadcrumb GROUP BY trip_id, DATE(tstamp)) AS results)) AS choice ON (breadcrumb.trip_id = choice.trip_id AND date(breadcrumb.tstamp) = choice.date);
2020-10-01
Trip ID 169302880
05:32:21



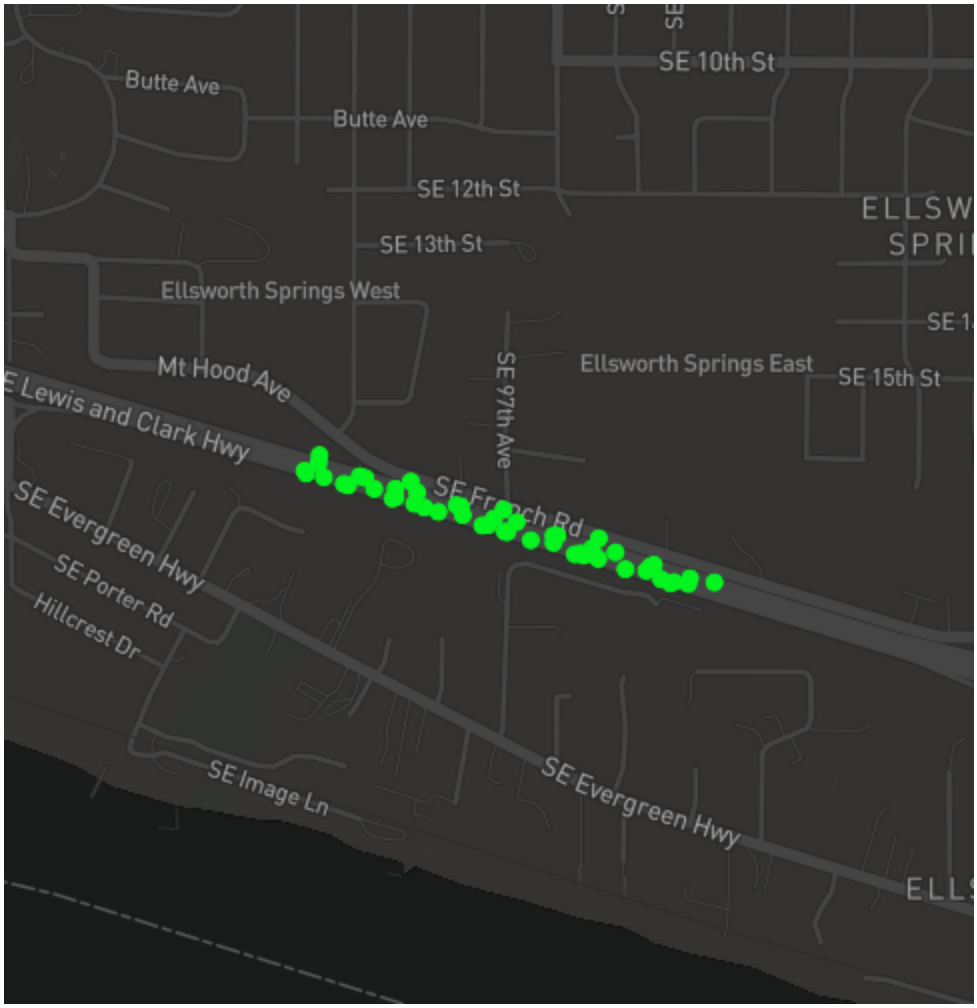
Bus speeds for the breadcrumb data of the longest trip (in time).This data was recovered from 1OCT2020 which had a trip_id of 169302880, and had a time of 05:32:21.

Visualization 5a, 5b, 5c, …. Three or more additional visualizations of your choice. Indicate why you chose each particular visualization.

1. List all bread crumb readings for a specific portion of Highway 14 (bounding box: [(45.610794, -122.576979), (45.606989, -122.569501)]) during
    a. Mondays between 4pm and 6pm.
    SELECT longitude, latitude, speed FROM breadcrumb WHERE EXTRACT(dow FROM DATE(tstamp)) = 1 AND (EXTRACT(hour FROM tstamp) >= 16 AND EXTRACT(hour FROM tstamp) <= 18) AND (longitude >= -122.576979 AND longitude <= -122.569501) AND (latitude >= 45.606989 AND latitude <= 45.610794);

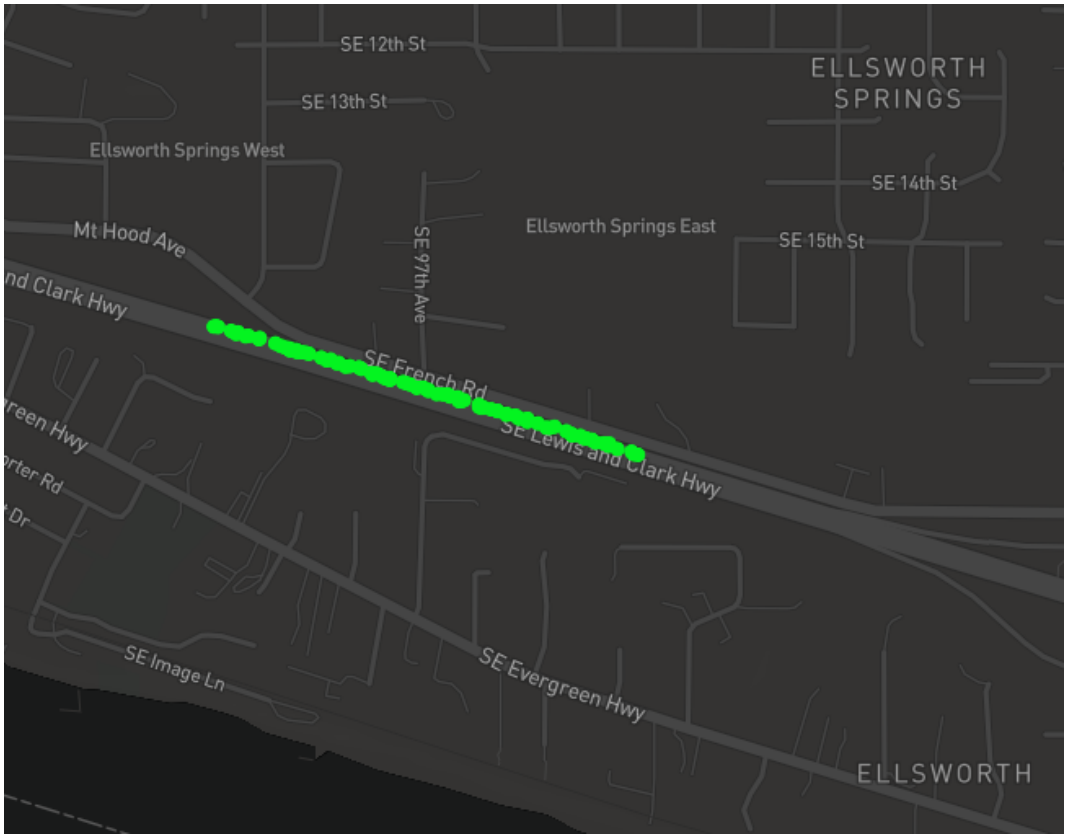
Bus speeds for the breadcrumb data pertaining to the largest trip_id that was within the region of [(45.610794, -122.576979), (45.606989, -122.569501)]) on Mondays from time frame of 1600-1800.

    b. Wednesday between 10am and 12pm
    SELECT * FROM breadcrumb WHERE EXTRACT(dow FROM DATE(tstamp)) = 3 AND (EXTRACT(hour FROM tstamp) >= 10 AND EXTRACT(hour FROM tstamp) <= 12) AND (longitude >= -122.576979 AND longitude <= -122.569501) AND (latitude >= 45.606989 AND latitude <= 45.610794);



Bus speeds for the breadcrumb data pertaining to the largest trip_id that was within the region of [(45.610794, -122.576979), (45.606989, -122.569501)]) on Wednesdays from time frame of 1000-1200.

c. Sundays between 6am and 8am.
SELECT * FROM breadcrumb WHERE EXTRACT(dow FROM DATE(tstamp)) = 0 AND (EXTRACT(hour FROM tstamp) >= 18 AND EXTRACT(hour FROM tstamp) <= 20) AND (longitude >= -122.576979 AND longitude <= -122.569501) AND (latitude >= 45.606989 AND latitude <= 45.610794);



Bus speeds for the breadcrumb data pertaining to the largest trip_id that was within the region of [(45.610794, -122.576979), (45.606989, -122.569501)]) on Sundays from time frame of 1800-2000.

I chose these three to show the difference between what would be between more busy times during the day and night times as well. I have Monday 1600 - 1800, Wednesday 1000 - 1200, and Sunday 1800 - 2000. For Monday, I chose this time frame so that it would be around the time of day when people would be attempting to go home. Wednesday would be for the middle of the day on a weekday, around lunch time. The Sunday mapping would be to demonstrate the evening traffic during the end of the weekend. As can be observed, the least amount of traffic would occur on the Wednesday option, while Monday's would be the time frame with the most readings.

Links for code

https://github.com/marobertson11/CS410/blob/main/gatherData.py
https://github.com/marobertson11/CS410/blob/main/data_producer.py
https://github.com/marobertson11/CS410/blob/main/data_consumer.py
https://github.com/marobertson11/CS410/blob/main/run.sh
https://github.com/marobertson11/CS410/blob/main/website-consumer.py
https://github.com/marobertson11/CS410/blob/main/website-producer.py