Jetha Harsha
Kaushik
Avinash
Harie Vashini

DataEng: Project Assignment 3

Data Integration

Submission

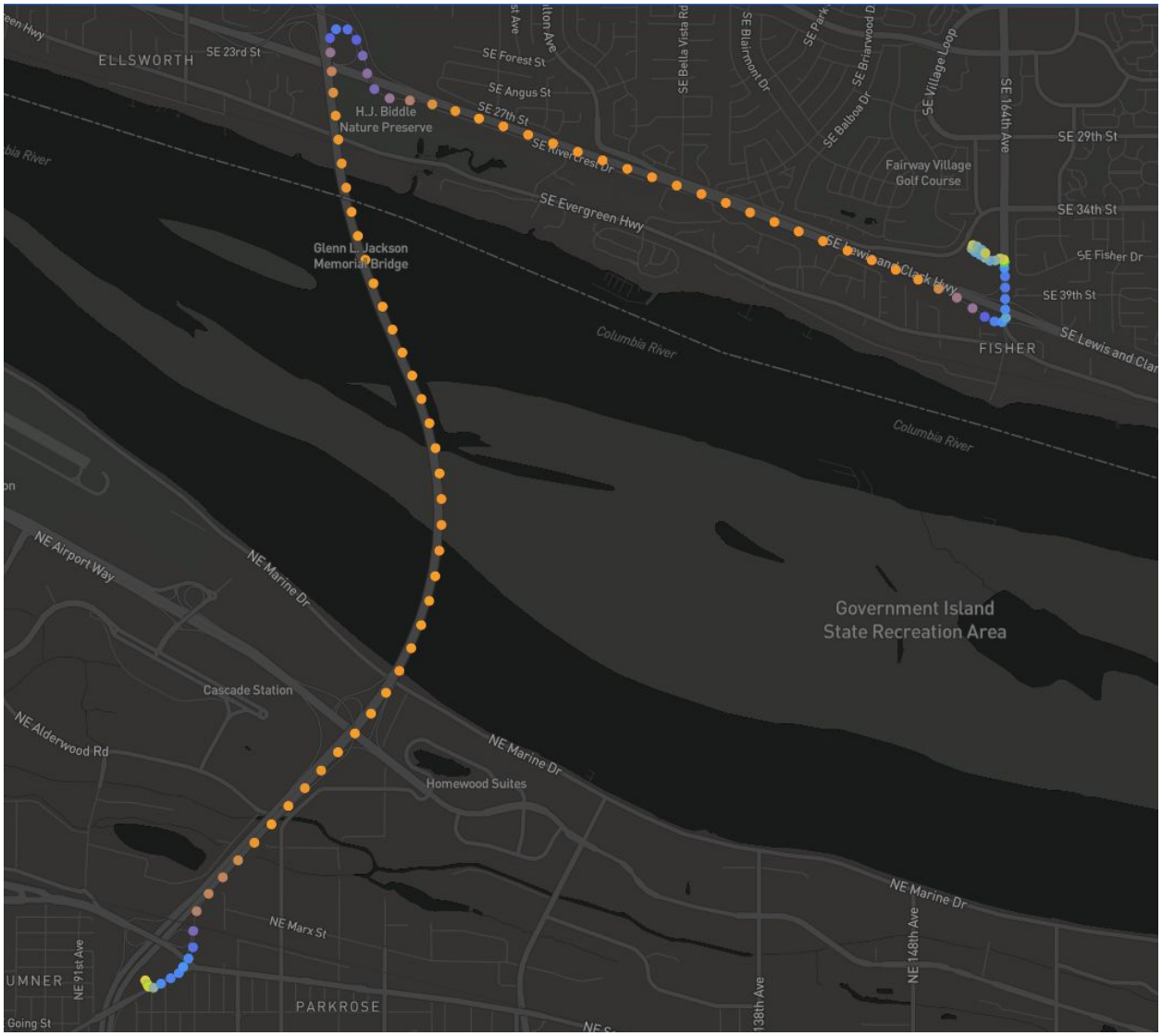
Make a copy of this document and update it to include the following visualizations. For each visualization extract from your database a list of {latitude, longitude, speed} tuples and then use the provided visualization code (see Section D above) to display bus speeds at all of the corresponding geographic coordinates. So, for example, if you are asked to visualize a “trip”, then you must query your database to find all of the {latitude, longitude, speed} tuples for that trip, and then display a map showing the recorded/calculated bus speed at each {latitude,longitude} location.

No need to produce software that neatly displays trips, routes, dates, times, etc. onto the visualization itself. Instead, just paste a screen capture of the map-based speed visualization into your submission document and then include a text description of the contents of the visualization. For example, text like this: “Bus Speeds for all outbound trips of route 65 between 9am and 11am on Sunday October 32, 2020.”

The default speed limits and the colour choices were not representative of the data we have. Our team believed that going forward with them would not depict the speeds of busses appropriately. For this reason, we have changed both the speed limits and their respective colours which convey what we wanted to show in a more practical way.

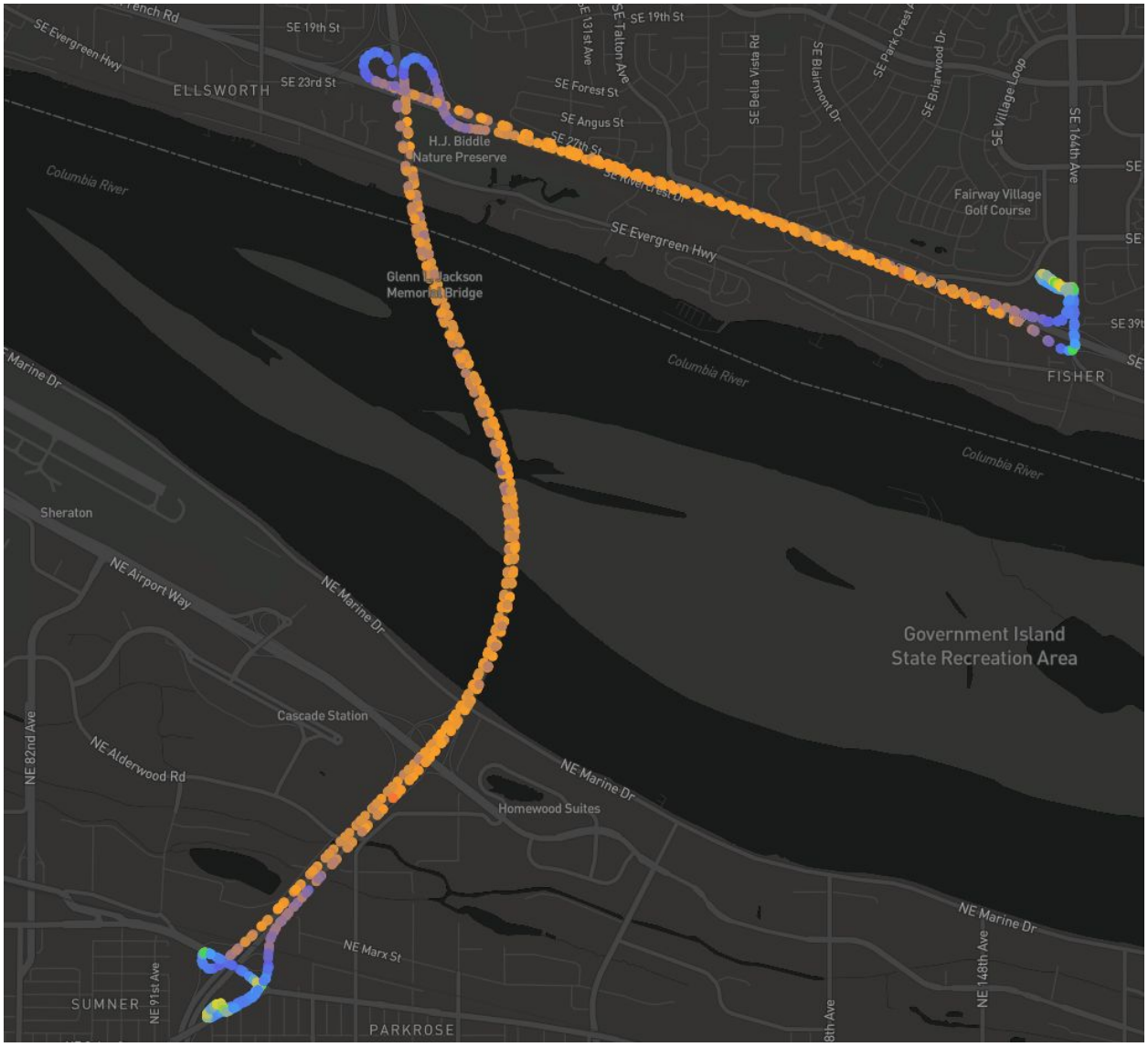
```
stops: [
            [0, 'rgb(50, 215, 75)'],  // Green
            [5, 'rgb(255, 214, 10)'], // Yellow
            [15, 'rgb(65, 156, 255)'],  // Blue
            [35, 'rgb(94, 92, 230)'],   // Indigo
            [60, 'rgb(255, 159, 10)'],  // Orange
            [80, 'rgb(255, 69, 58)']  // Red
        ]
```

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!



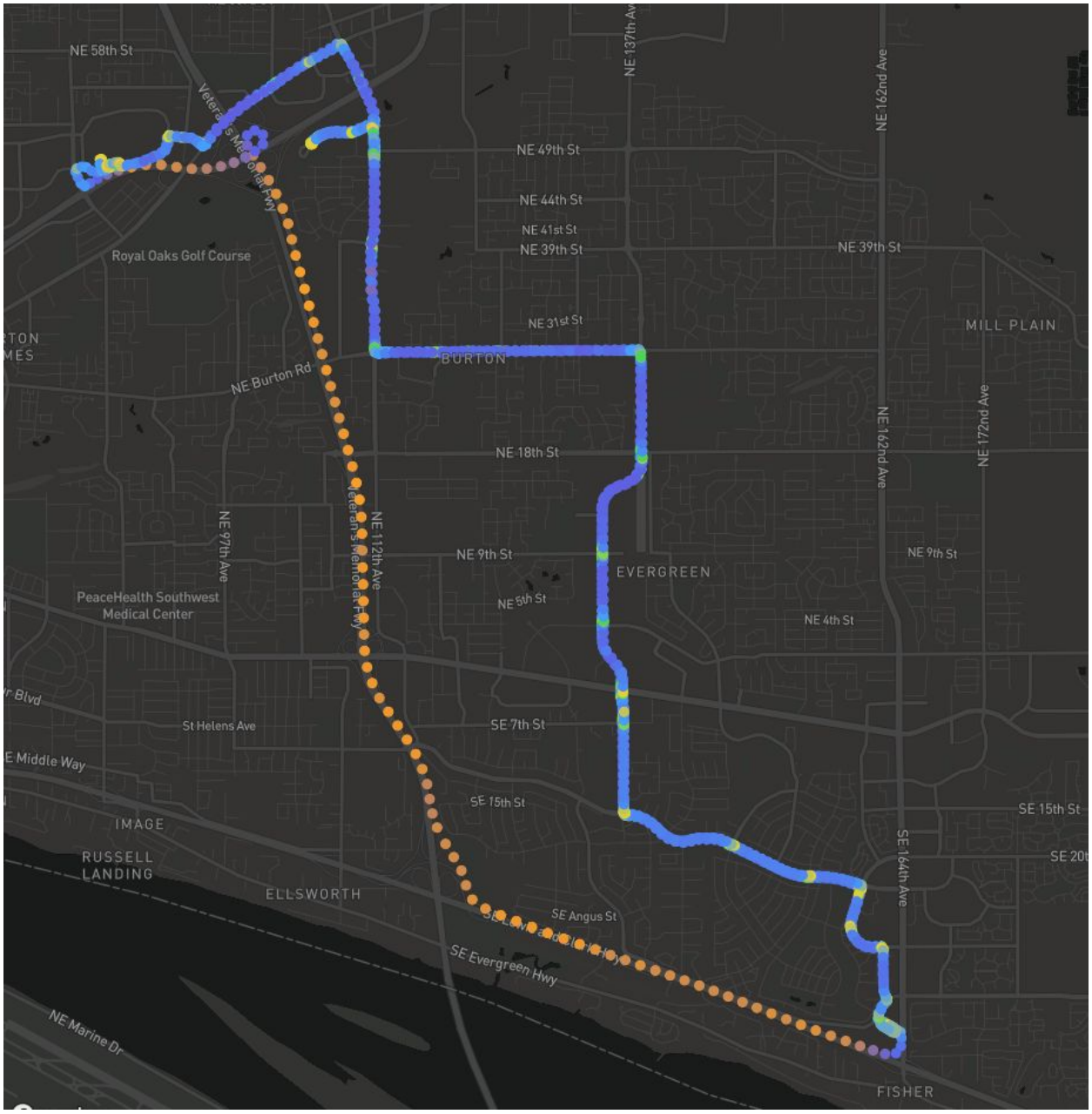
Visualization 2. All outbound trips that occurred on route 65 on any Sunday (you choose which Sunday) between the hours of 4pm and 6pm.

Visualization 3. All outbound trips for route 65 on any Sunday morning (you choose which Sunday) between 9am and 11am.



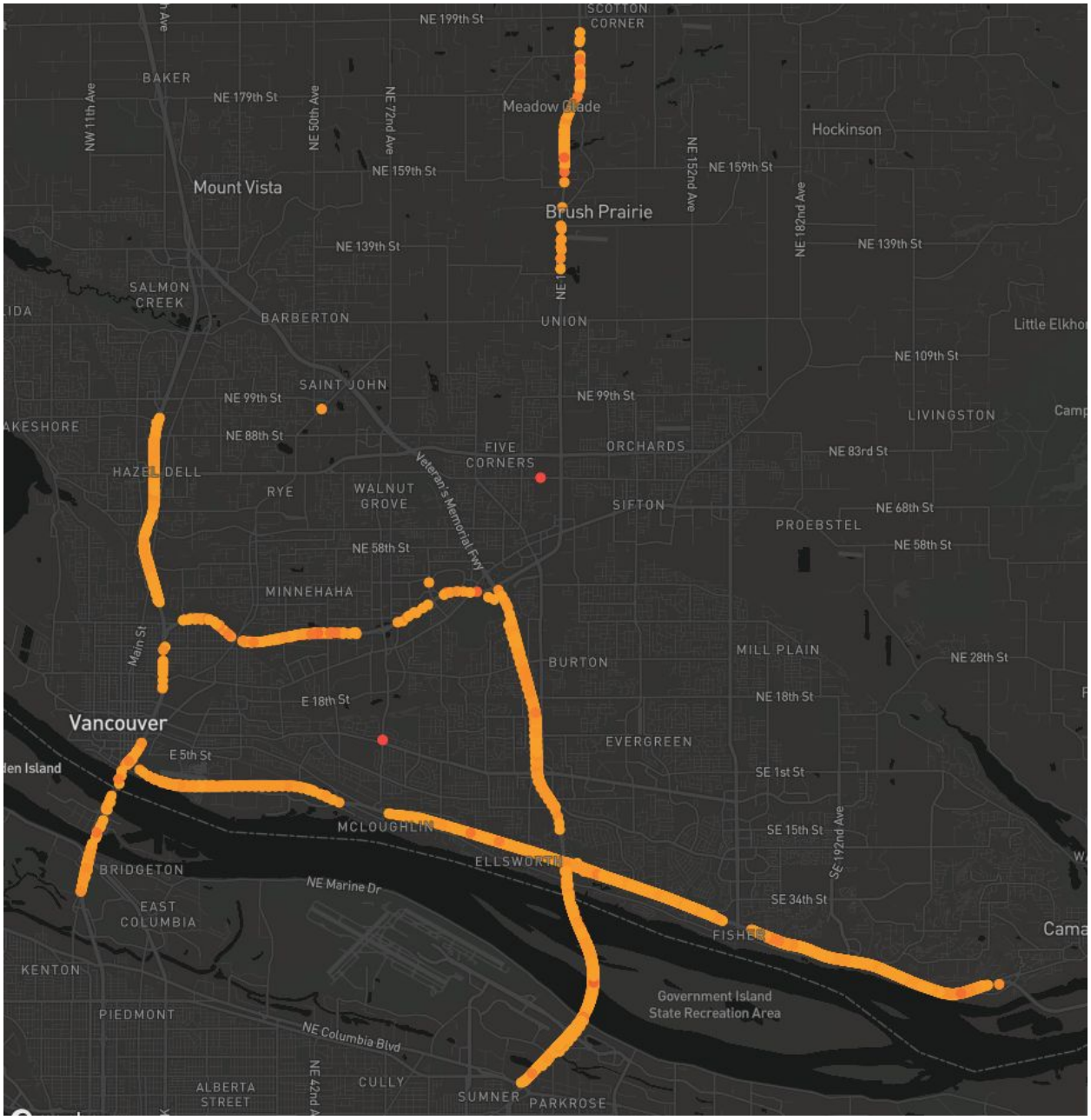
Visualization 4. The longest (as measured by time) trip in your entire data set. Indicate the date, route #, and trip ID of the trip along with a visualization showing the entire trip.

Date - 2020-09-29
Route - No route Id because there is no route details for that date (Started recording from 10th Oct, 2020)
Trip_ID - 169026975
Longest trip time - 03:12:05

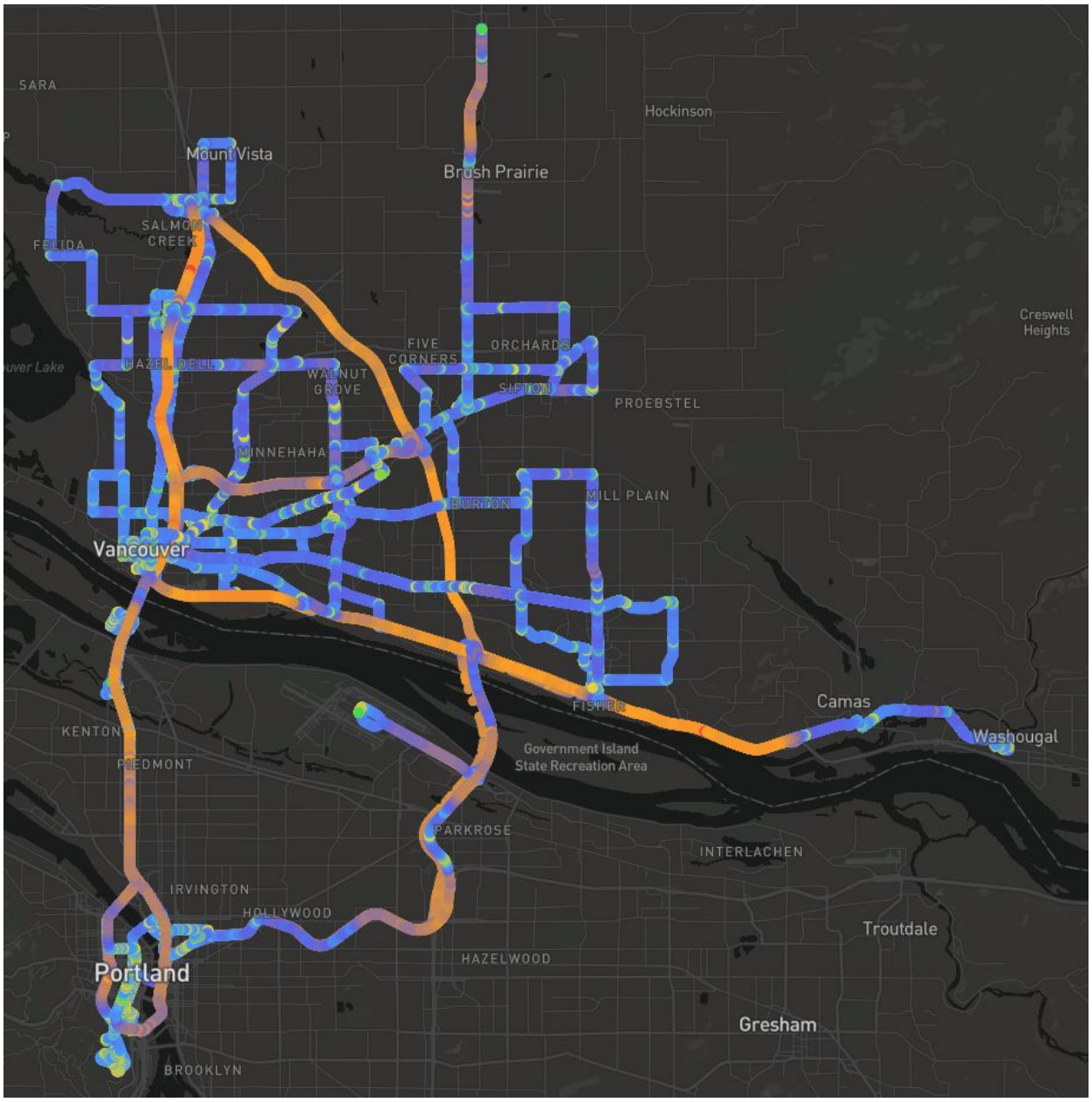


Visualization 5a, 5b, 5c, …. Three or more additional visualizations of your choice. Indicate why you chose each particular visualization.

5a. We believe public transport should not go above 60 mph. So this graph shows at what locations the buses are going above 60 mph. This information can be used to put more signs to reduce the speed.



5b. Curious to see the operations of all the buses in one day? We were too. So we choose to see all the operations on 4 Sep 2020.



5c. Which route was taken the highest by all the busses. We are curious to see which route is the busiest.

Queries

1. \COPY (select latitude,longitude,speed from breadcrumb where trip_id='167236967') TO 'DE/examples/clients/cloud/python/map_files/query1.tsv' WITH NULL AS '' DELIMITER E'\t' CSV HEADER;

2. \COPY (select B.latitude, B.longitude, B.speed from breadcrumb2 B, trip2 T where date(B.tstamp) = '2020-10-11' and B.trip_id=T.trip_id and T.route_id=65 AND cast(B.tstamp AS time) between '16:00:00' and '18:00:00')TO 'DE/examples/clients/cloud/python/map_files/query2.tsv' WITH NULL AS '' DELIMITER E'\t' CSV HEADER;

3. \COPY (select B.latitude, B.longitude, B.speed from breadcrumb2 B, trip2 T where date(B.tstamp) = '2020-10-11' and B.trip_id=T.trip_id and T.route_id=65 AND cast(B.tstamp AS time) between '09:00:00' and '11:00:00')TO 'DE/examples/clients/cloud/python/map_files/query3.tsv' WITH NULL AS '' DELIMITER E'\t' CSV HEADER;

4. To find the trip_id with the longest trip time we wrote the query as:
   select age(max(tstamp),min(tstamp)) as time, trip_id from breadcrumb2 group by trip_id order by time desc LIMIT 1;

   From the above founded trip_id we wrote the following query to visualize:
   \COPY (select B.latitude, B.longitude, B.speed from breadcrumb2 B, trip2 T where B.trip_id=T.trip_id and T.trip_id=169026975)TO 'DE/examples/clients/cloud/python/map_files/query4.tsv' WITH NULL AS '' DELIMITER E'\t' CSV HEADER;

5. a) \COPY (SELECT B.latitude, B.longitude, B.speed FROM breadcrumb2 B, trip2 T WHERE T.service_key IN ('Saturday','Sunday') AND B.speed > 60 AND B.trip_id = T.trip_id)TO 'DE/examples/clients/cloud/python/map_files/query5a.tsv' WITH NULL AS '' DELIMITER E'\t' CSV HEADER;

   b) \COPY (SELECT latitude, longitude, speed FROM breadcrumb WHERE (date(tstamp) = '2020-09-04' AND cast(tstamp as time) BETWEEN '04:00:00' AND '23:59:59') OR (date(tstamp) = '2020-09-05'AND cast(tstamp as time) BETWEEN '00:00:00' AND '02:00:00'))TO 'DE/examples/clients/cloud/python/map_files/query5b.tsv' WITH NULL AS '' DELIMITER E'\t' CSV HEADER;

   c) \COPY (SELECT B.latitude, B.longitude, B.speed FROM breadcrumb2 B, trip2 T WHERE B.trip_id= T.trip_id AND T.route_id IN (SELECT T.route_id FROM breadcrumb2 B, trip2 T WHERE B.trip_id= T.trip_id and T.route_id is not null GROUP BY T.route_id ORDER BY COUNT(*) DESC LIMIT 1))TO 'DE/examples/clients/cloud/python/map_files/query5c.tsv' WITH NULL AS '' DELIMITER E'\t' CSV HEADER;