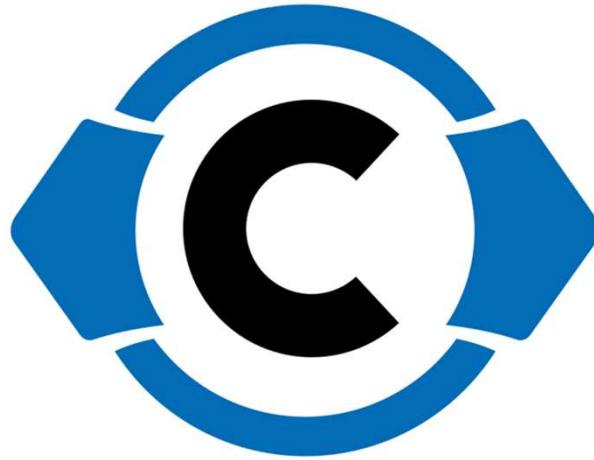


Smart City: StopSpot



C-TRAN

Brian Allen, Marcus Kwong, Tarun Singh, Hui-Yu Sim

C-Tran's Goal and Minimum Viable Product

Goal:

- The goal of C-Tran is to improve the accuracy of static GPS positioning data of C-Tran transit stop using dynamic vehicle location data.

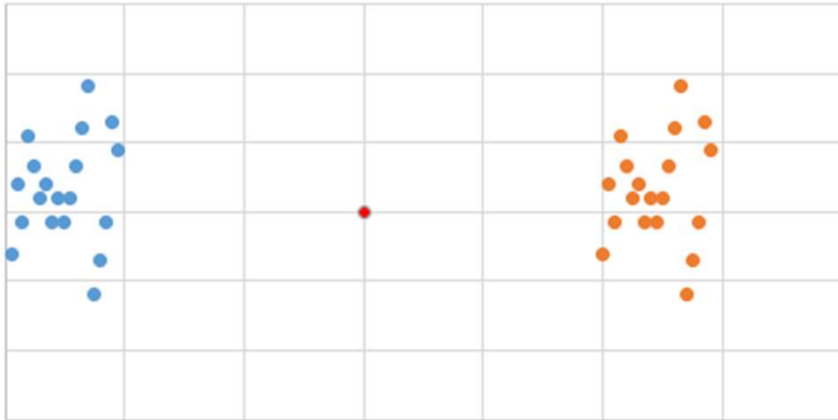
MVP:

- A list ranking from worst to best accurate stops.
- Stretch goal
 - a. Visualization tools



Analysis

1. What count as an accurate stop?
2. Finding percent error of each stop location using the dynamic vehicle location data
3. Each one of us did our own analysis and implementation
 - a. Data cleaning
 - b. Creating visualization



End of two-week Sprint

1. Successfully deliver the list
2. Demo our visualization



Preprocessing : Decision Making

- Restricting Attributes
 - Removal of excess attributes
- Removal of Outliers
 - Null X & Y Coordinates
 - Multi-stops
- Unscheduled/Temporary Stops
 - Schedule Status values 0, 1, 3

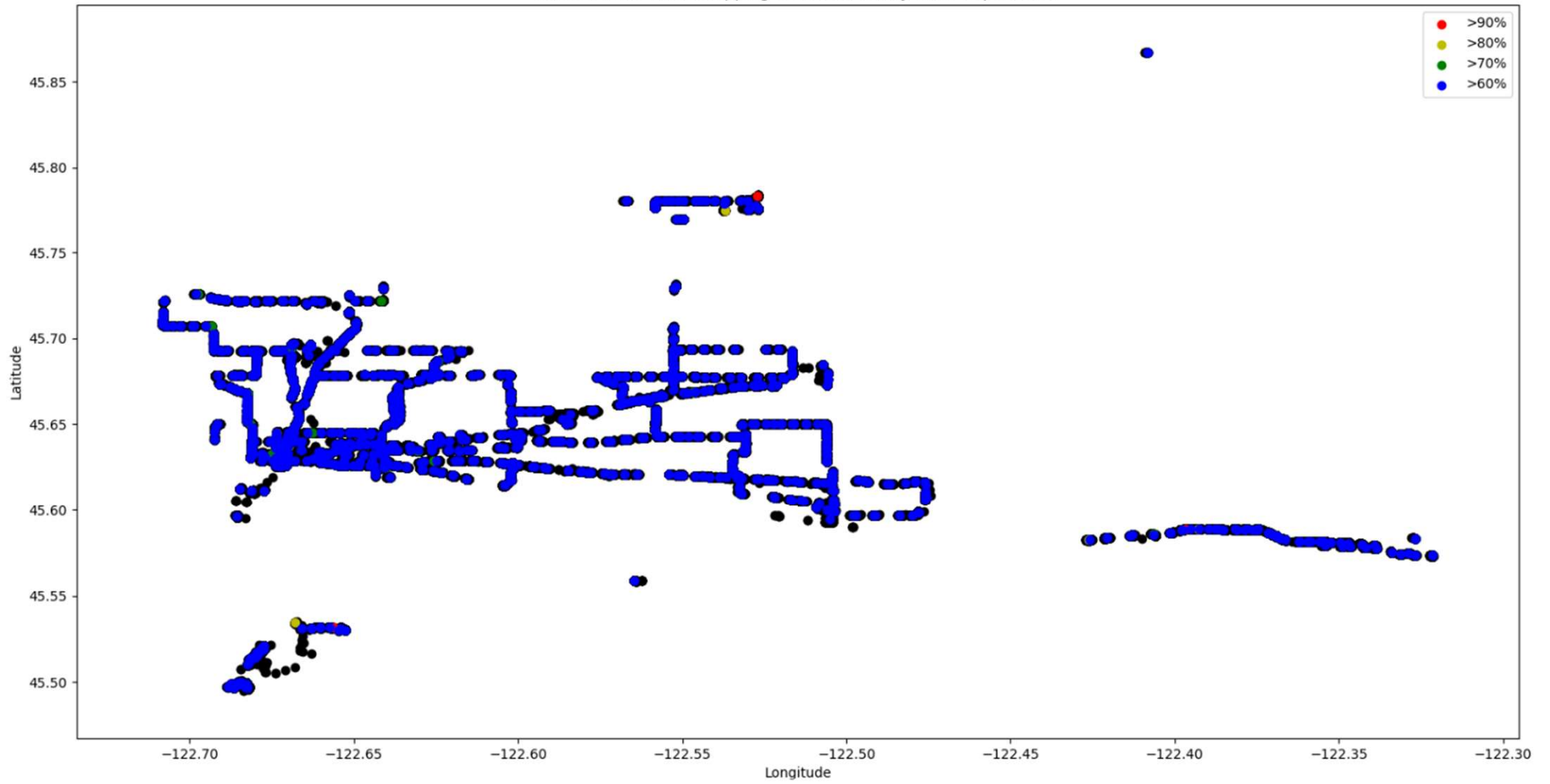


The Results: Tables and Graphs

- Preprocessing Data for Analysis
- Produce Results through SQL query
- Use .csv file from SQL results for Visualization data

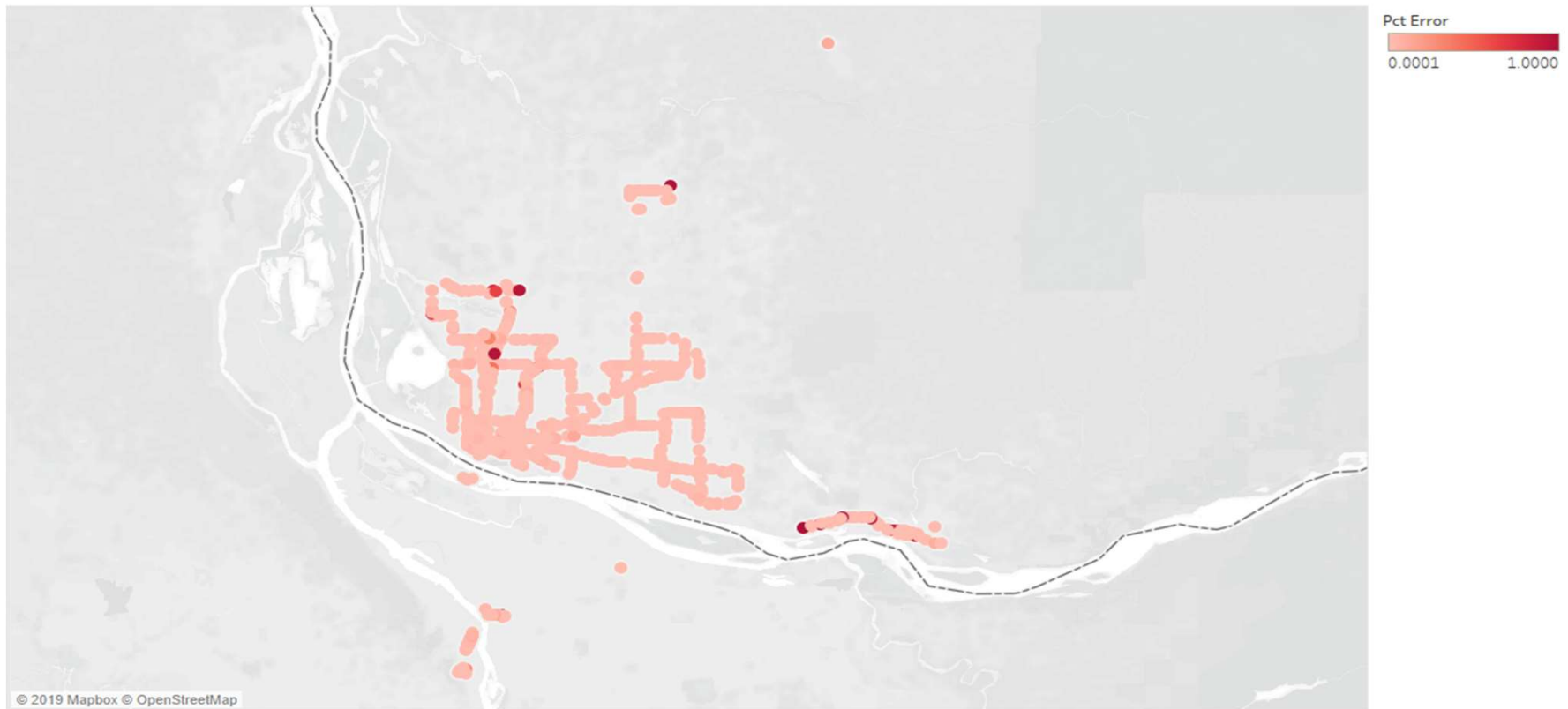
location_id	stop_name	stop_lon	stop_lat	total	greater_98	average_distance	pct_error
3179	NE Multnomah & 9th	-122.656174	45.531506	486	469	160.5611822	0.965
6038	Fairgrounds Ave at B	-122.52713	45.783257	313	296	458.6869326	0.9457
2706	NW 78th St & Fruit V	-122.691498	45.6786	1223	1156	176.9271059	0.9452
1169	NE 3rd Ave 1400 Blc	-122.396698	45.588688	2657	2483	603.7526463	0.9345
544	Columbia & 33rd St	-122.673508	45.645229	1252	1079	124.89245	0.8618
6070	Evergreen Blvd & Co	-122.673981	45.628769	2562	2193	192.8285596	0.856
6040	Rasmussen & Parkwa	-122.536919	45.775002	2577	2167	126.6327728	0.8409
2672	Columbia & Fourth F	-122.673698	45.640671	1258	1056	140.4847796	0.8394
533	Columbia & McLoug	-122.673698	45.635288	1260	1037	180.5492339	0.823
3380	NE 3rd Ave & 1st Ave	-122.392754	45.588951	2700	2195	167.0210789	0.813
3701	NE 139th St & 29th	-122.641579	45.721851	1992	1617	142.2864619	0.8117
964	NW 6th Ave & Nonwo	-122.426132	45.58239	2685	2173	134.938148	0.8093
6133	Columbia & 15th St	-122.673859	45.63269	1253	1007	180.2960783	0.8037
4143	Hazel Dell Ave & 10	-122.668983	45.694321	1198	962	191.9962594	0.803
6060	Fourth Plain & Danie	-122.675148	45.640152	2777	2223	194.7616477	0.8005

Percent error of bus stopping above 98ft away from stop location



Percent error from 0.0001 to 1.0000 for all stops.

Sheet 1



Map based on Stop Lon and Stop Lat. Color shows details about Pct Error. Details are shown for Location Id, Stop Name and Total. The view is filtered on Pct Error, which ranges from 0.0001 to 1.0000.

Percent error above 85%

Sheet 1



Map based on Stop Lon and Stop Lat. Color shows details about Pct Error. Details are shown for Location Id, Stop Name and Total. The view is filtered on Pct Error, which ranges from 0.8020 to 1.0000.

Filtering above 85% error with number of bus stopping iterations.

Sheet 1



Map based on Stop Lon and Stop Lat. Color shows details about Pct Error. Size shows details about Total. Details are shown for Location Id and Stop Name. The view is filtered on Pct Error > 0.8571.

Data with filtering of 85% above error and number of bus stopping iterations.

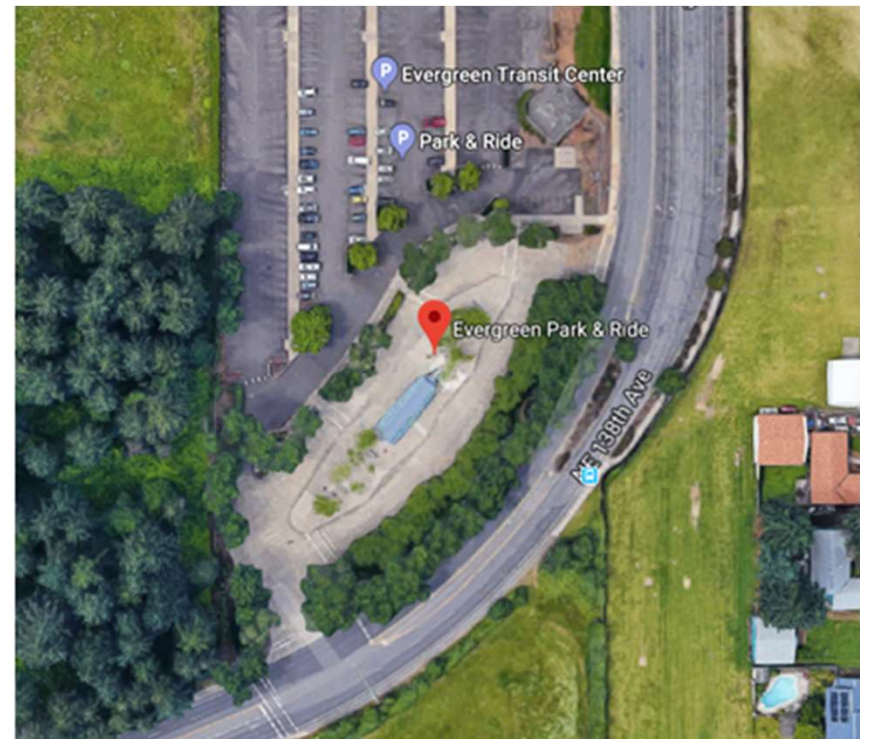
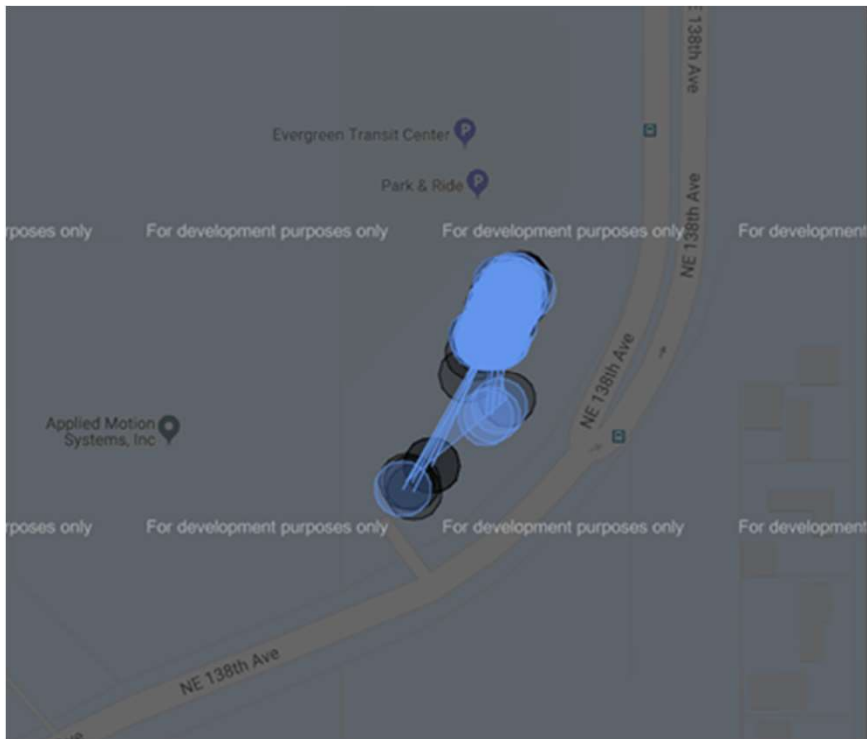
☰ View Data: Sheet 1

Show aliases

Location Id	Pct Error	Stop Lat	Stop Lon	Stop Name	Total
1336	0.959000	45.684727	-122.660873	Hwy 99 8500 Block	1,586
2351	0.958300	45.577980	-122.341881	Main & 28th St	1,509
3179	0.997800	45.531506	-122.656174	NE Multnomah & 9th	461
2159	1.000000	45.584888	-122.413544	NW 6th Ave 400 Block [eb]	456
2329	0.997100	45.581249	-122.358612	E St 1200 Block	347
3037	0.964700	45.708488	-122.708252	NW 36th Ave & 122nd St	170
1169	1.000000	45.588688	-122.396698	NE 3rd Ave 1400 Block	127
2321	1.000000	45.588360	-122.375023	NE 3rd Ave & Shepherd Rd	123
6038	0.986500	45.783257	-122.527130	Fairgrounds Ave at Ball Park	74
3701	0.950800	45.721851	-122.641579	NE 139th St & 29th Ave [eb]	61
2079	0.916700	45.676952	-122.626663	St Johns & 43rd Ave [nb]	24
4174	0.857100	45.692951	-122.680038	NW 99th St & 9th Ave [wb]	7
964	1.000000	45.582390	-122.426132	NW 6th Ave & Norwood	1

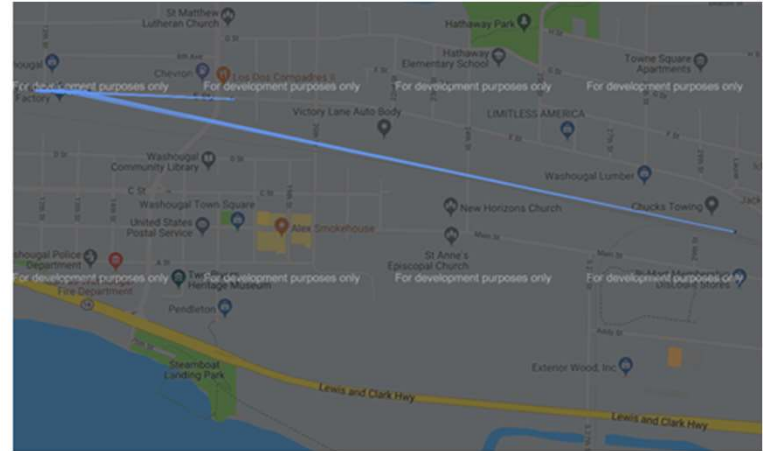
Graphical Results: The Good

Stop 102

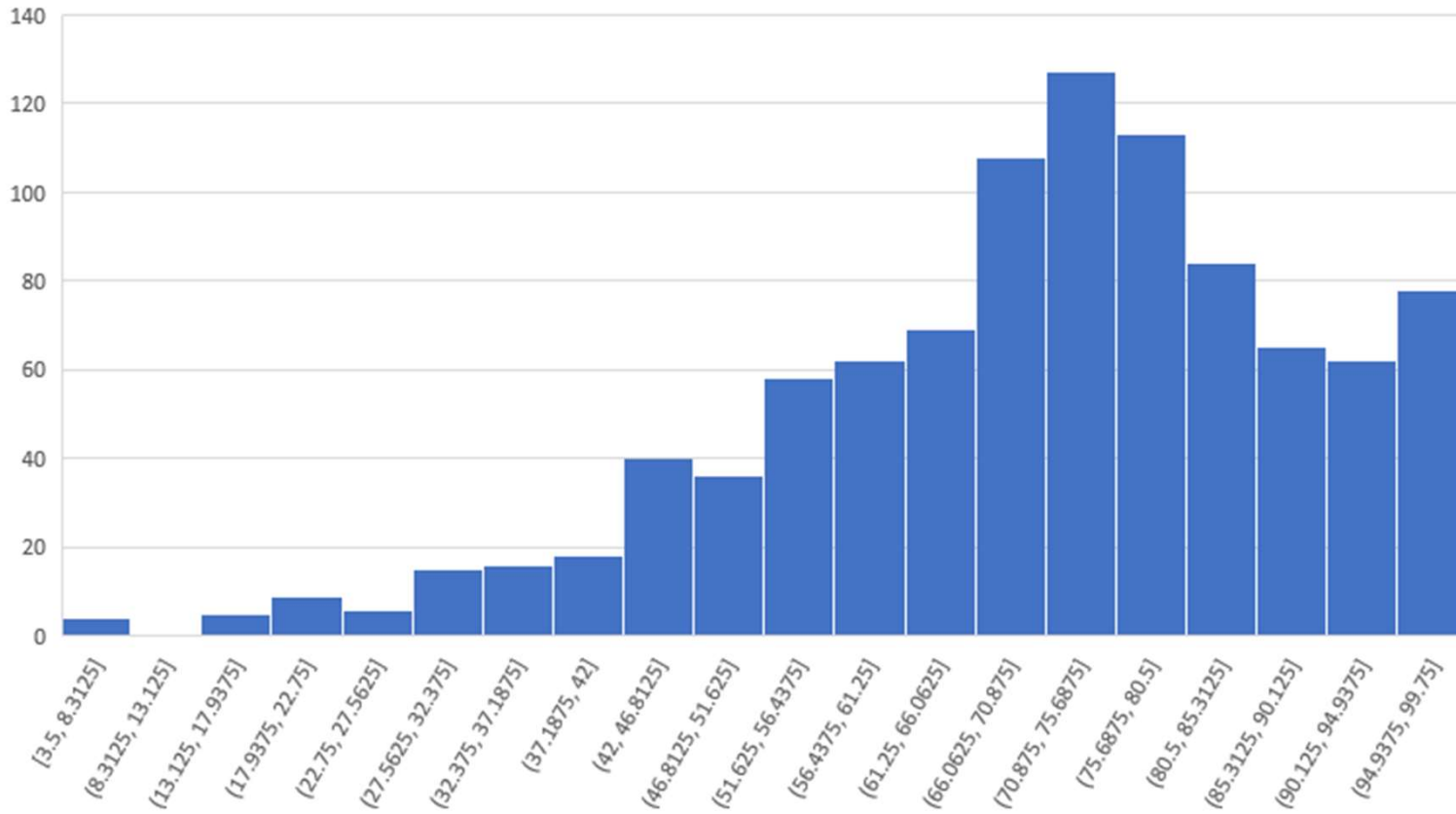


Graphical Results: The Bad

Stop 2329



Accuracy Metrics of Bus Stop Instances within 98 Feet



Data Wrangling

Trying to get the data

From PGE to C-Tran

Getting the actual data

The screenshot shows an email inbox with the search filter 'label:data-sci'. The messages are as follows:

- Hui, me 2**: Final presentation link - Thanks man. I'll section in awhile. On Sun, Aug 11, 2019, 2:43 PM Hui Yu Sim <hui2@pdx.edu> wrote: hey all, I've been working on the powerpoint slides in our share...
Attachments: Final Presentati...
- Kristin Tufte**: Invitation: Reset Invite: Portal Users Group @ Wed Aug 21, 2019 2pm - 3:30pm (PDT) (briallen@pdx.edu) - You have been invited to the following event. more details » Reset Invite: Portal U...
- Kristin Tufte**: PORTAL PUG Meeting Wed 8/21 - High-Resolution Transit Data - Good Morning Everyone - This email is a quick reminder of the August Portal Users Group Meeting - next Wednesday, Augu...
- me, Marcus 7**: Meet up? - Heh, totally misunderstood. I'll see you Tuesday at 930 ish. On Mon, Aug 12, 2019, 12:09 AM Marcus Kwong <mkwong@pdx.edu> wrote: Don't we have class at noon? Did you m...
- me, Hui 4**: Data Sci outline - Rad. Thank you. Looking forward to getting this presentation done. On Sun, Aug 11, 2019 at 1:36 PM Hui Yu Sim <hui2@pdx.edu> wrote: Some of the stuff that we learned...
- Bruce, Marcus 2**: reminder: stopspot demo meeting today - Hello, Here is a link to our results data we will be presenting in the meeting. Please feel free to look over the documents and we will be more than ...
Attachments: Deliverables for...
- Hui .. Marcus, me 28**: C-Tran analysis - Thanks, man. Ran out of steam tonight. Waking up early to look everything over.
Attachments: 4123_stops.csv, 4123.csv
- me .. Marcus, Hui 10**: Project ideas. - updated the github with your changes. Realized I didn't sent this yesterday. Sorry about that.
- me .. Tarun, Marcus 4**: Meeting with Kristin - That works for me as well
- Kristin Tufte**: Accepted: C-Tran Group Midway Meeting @ Thu Aug 1, 2019 2:10pm - 2:40pm (PDT) (briallen@pdx.edu) - Kristin Tufte has accepted this invitation.C-Tran Group Midway MeetingWould hav...
- Kristin Tufte**: Invitation: stop-spot-ctran @ Thu Jul 25, 2019 2:30pm - 3:30pm (PDT) (briallen@pdx.edu) - You have been invited to the following event. more details » stop-spot-ctranKristin Tufte is invitin...
- me, Hui 10**: Proposal Outline - Wow. Looks great. Thank for putting together the timeline and milestones - five weeks will go by so fast.
Attachments: Data Sci Project, Proposal Outline
- hui2018**: hui2018 invited you to hui2018/SmartCity_StopSpot - @hui2018 has invited you to collaborate on the hui2018/SmartCity_StopSpot repository You can accept or decline this invitation. You ...
- Bruce Irvin (via Go)**: StopSpot - invitation to collaborate - rbi@pdx.edu has invited you to contribute to the following shared folder:StopSpot OpenGoogle Drive: Have all your files within reach from any device. G...
- Kristin .. me, Bruce 6**: Fwd: c-tran data - Great, thanks Brian and Hui Yu. Yes, let's talk at class tomorrow. Just now I added Hui Yu, Brian and Tarun to the slack channel ds-stopspot-project
- Kristin Tufte**: Out of the Office Re: c-tran data - Hello - I am out of the office and will return Thursday the 18th. I will have limited email access and will respond to most messages when I return. Thank yo...
- me, Marcus 2**: Data Science Project - Hi, I would also agree to move forward with this project. I think we should also touch base with the other student who had contacted Bruce about this project.

Analysis Reflection

Writing queries vs writing code

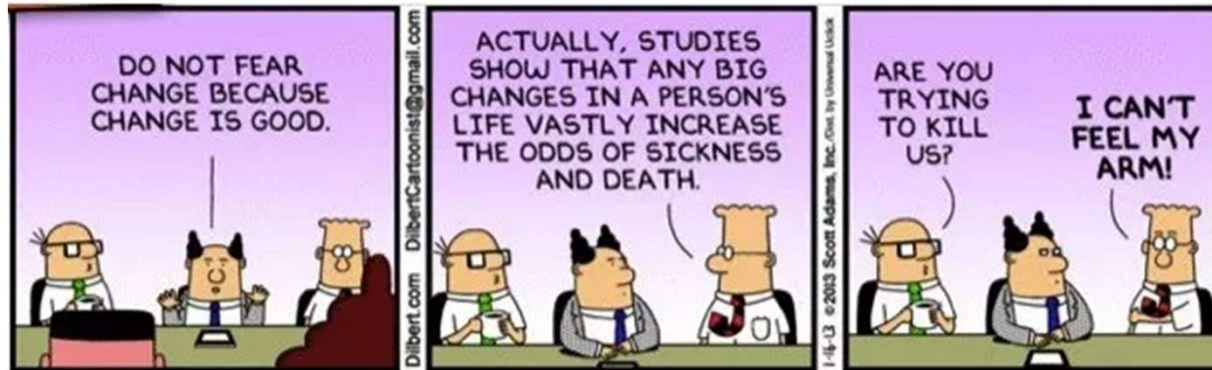
Statistical analysis

Lazy spreadsheets

What do you mean we have different data?



What we would change?



What would we change?

Our tech stack

Cleaning process

Better questions leads to better answers

Read the documentation then read it again.



Backlog

UI

Create a front end that can run a set of queries

- Show highest percent error
 - At X ft away
 - Remove stops X away
 - With door flag on or off
 - Remove stops with less than X occurrences

Add “sliders” for accuracy of stops (David ask)

- (Show all \longleftrightarrow Only show highest error)

Create a front end that displays maps pins for a single stop

Create a front end that displays pins for all stops

- Create a front end that can click on a stop to get single stop map

Create a front end that can be fed quarterly data and stop info for analysis

Backend

Choose and setup database

Choose framework to use to serve pages

- Flask?
- Django?
- ?

What Map package to use?



Backlog (cont)

Science

Looks for vehicles with consistent GPS issues
Double check accuracy of location_distance
Investigate geographic impacts on gps accuracy

Technical Debt

Create shared work space for Marcus and I

- Docker on the cloud? PSU space?

New github?

Other

Create a mastery query that can clean and analyze data is one go

- Just use python for analysis?

Flag for checked or accurate stops

