UBC Sustainability Scholars Program
Metro Vancouver, Air Quality and Climate Change

Metro Vancouver Smart Drive Challenge:
"Can We Drive Less?"

Prepared by: Lena Patsa
PhD Candidate (UBC Mining Engineering)
UBC Sustainability Scholar, 2017

Prepared for: Conor Reynolds, Ph.D., P.Eng.
Program Manager, Air Quality & Climate Change Policy
Parks, Planning and Environment Department

The University of British Columbia
September 2017

Table of Contents

List of Figures

List of Tables

1 Executive Summary

The Smart Drive Challenge Project was conducted between October 2016 and January 2017 and involved 200 drivers from across the Metro Vancouver Regional District (MVRD). This innovative pilot study used on-board location tracking and "smart" connected feedback technology to determine whether short-term driver training and "smart" feedback can help improve fuel consumption and fuel efficiency, and instil more environmentally-friendly driving habits.

Metro Vancouver hired a UBC Sustainability Scholar to help with the analysis and interpretation of the large amount of GPS and participant-survey data that was collected at the completion of the Smart Drive Challenge Project. The overall objective of the analysis focused on determining whether the training and feedback helped drivers change their driving habits and behaviour, and drive less. The study scope included additional exploratory analysis, using a variety of techniques, to 1) assess and improve upon the extent and quality of the collected data; and 2) to improve upon the overall value of the collected data by identifying potential opportunities for further study.

The original data was provided as a large collection of spreadsheets and text files that in their original form were not suitable for extensive analysis. The data was cleaned up, re-structured, and processed, in order to transform it into a more dynamic and flexible form that would be more amenable to different types of analytical approach. Preliminary statistical analysis identified and allowed for the removal of low-quality entries (such as empty or mis-calibrated records). The resulting database stores GPS locations and associated qualifying attribute information for all participating drivers.

The scrubbed and transformed data was subsequently examined in different ways, namely: statistically (using MS Excel and R) to describe it; geographically (using ArcGIS) to visualize the physical extent of all recorded paths, and to identify any geographical outliers; and visually (using Tableau) to examine the routes of each driver, and to visually identify shifts in emerging pre- and post- driving patterns.

Initially, results indicated very small post-training improvements, mainly in fuel efficiency. Upon a more detailed inspection and following the geographical analysis, which revealed that a portion of the data belonged to longer inter-city travel (that in some cases extended all the way to Calgary, Alberta), the design decision was made to remove longer trips from the set and focus the analysis on trips that did not transverse a modified version of the MVRD boundary. This decision was further supported by the fact that the main focus of the study was to determine the effects of the training on changes in inner-city driving (most specifically within the MVRD), and not cross-country travel. Subsequent analysis of the subset data indicated significant changes in post-training driving behaviour, the most significant of which was a reduction in the daily average hard accelerations and decelerations (by -36.05% and -31.14% respectively), as well as in the hard-to-total ratios of accelerations and decelerations (by -31.42% and -27.61% respectively). The overall values of period-averaged km-distance travelled (-4.9%), average trip speed (-5.75%), and average number of completed trips (-5.8%) also dropped. There was also a smaller overall reduction in the period-averaged driving (-2.81%) time, the period-averaged idling time (-1.23%), and the idling fuel consumption (-3.38%).

2 Introduction

Background

The Smart Drive Challenge Project was a pilot study conducted by MetroVancouver between October 2016 and January 2017. It used on-board vehicle location tracking and "smart" connected feedback technology to determine whether short-term driver training and "smart" feedback can help improve fuel consumption and fuel efficiency, and install more environmentally-friendly driving habits. The 200 volunteer drivers participating in the study received training that aimed to instil habits that would ideally help these drivers driver better and drive less.

Training consisted of an hour-long online course taken by all participants, and for a subset of drivers, an additional hour-long face-to-face training with a driver instructor. To establish a benchmark for comparison, driving activity was initially tracked for a 3-week period prior to training, termed *Baseline*. A week-long *Training* period followed, during which drivers attended the course (online and in-person) and were subsequently provided with "smart" feedback via email that would inform them of their progress as they undertook the final, monitoring stage of the study, termed *Challenge*. A subset of participants (termed the *Extended* group) continued on with the Challenge part of the study for a longer period of time that varied by driver but was at most an additional 7.5 weeks on-top of the 3-week long Challenge completed by the *Core* group of participants (Figure 1).

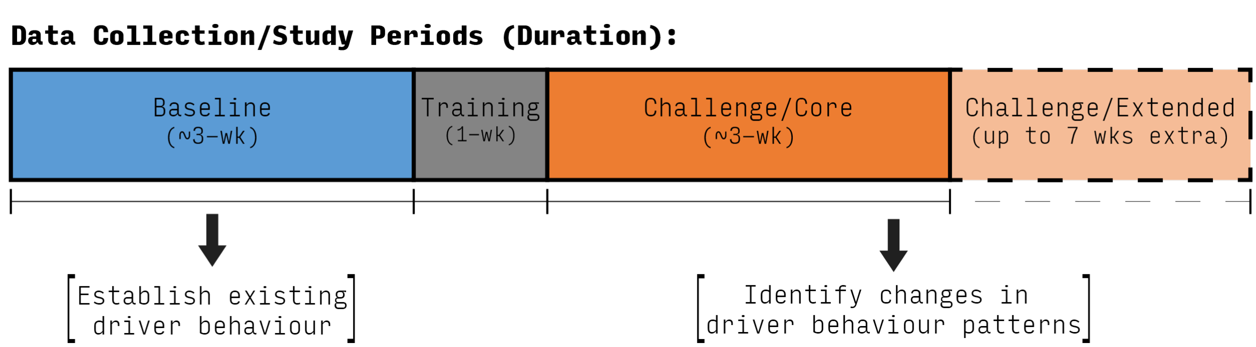

Figure 1: *Baseline, Training, and Challenge (Core & Extended) periods in the Smart Drive Challenge Project.*

Project Definition

Metro Vancouver hired a UBC Sustainability Scholar help with the analysis and interpretation of the large amount of GPS and participant-survey data collected during the Smart Drive Challenge Project.

The initial aim of the data analysis was to determine whether the training and on-board feedback helped drivers change their driving habits and behaviour, and drive less. Subsequently, the study scope was expanded to include additional exploratory analysis, using a variety of techniques, to: 1) assess and improve upon the extent and quality of the collected data; and 2) improve upon the overall value of the collected data by identifying potential opportunities for further study.

The overall objective was to evaluate, qualitatively and quantitatively, the impact of the provided training (on-line and in-person) and feedback technology on driver trip-making behaviour and on vehicle fuel consumption & greenhouse gas (GHG) emissions, using participant demographics and other data characteristics, in order to hypothesize as to if any driver sub-groups were (less or more) likely to change their driving behaviour.

Findings will potentially be used to inform future program development on: improving fuel efficiency and reducing emissions from light duty vehicles in the region. It also has the potential to affect any further studies of a similar nature, e.g. by identifying what worked and what didn't, or what other types of data should be collected. Study findings would also be of interest to stakeholders of future public education and social marketing campaigns (related to fuel consumption and GHG emissions), and those charged with issuing driving train-the-trainer recommendations.

3 Approach

Input Data

The data collected by the onboard vehicle tracking devices consists of timestamped GPS locations (i.e. timestamp and latitude/longitude pairs). Data gathered during the Baseline stage would help establish any pre-existing driving behavioural patterns, while data collected during the Challenge stage would ideally help identify any changes in driving behaviour as a result of the training drivers undertook half-way through the study period. Apart from the timestamped GPS locations, a series of driver surveys were completed to collect additional contextual information about the participants (covering mostly demographic and lifestyle information). These attributes were used in the subsequent analysis to add contextual context and to define driver subgroup of particular interest.

The data was made available to the Scholars as a collection of spreadsheets and .CSV text files. The spreadsheets consisted results of 3 surveys (Application, Baseline, and post-Challenge), a participant tracking spreadsheet, and a summary spreadsheet with individual trip data. More specifically:

— The Application Survey held data pertaining to driver and vehicle characteristics (such as driver age and driver income level, and vehicle make and model) was the main source of contextual data for the analysis.

— The Participant Tracking spreadsheet provided information relating to driver participation (such as the start and end dates of each driver's Baseline, Training, and Challenge periods), as well as participant-averaged values of basic variables of interest (such km-travelled, duration of travel, accelerations and deceleration events (hard and total), and fuel economy/consumption). This data was used to assign individual trips and their corresponding route increment points into one of 3 study periods (i.e. Baseline, Training, or Challenge).

- The Baseline Survey contained survey responses on various aspects of the participant's decision making and use of private and public transportation means (e.g. whether they typically took public transit to work, or rode a bike to University).
- The Post-Challenge Survey listed self-reported changes in driver behaviour during the study (e.g. switching to public transportation rather than driving), as well as intended use of public and private means of transportation after the study.
- An additional spreadsheet (Training Survey) containing qualitative participant feedback that was collected at the end of the training session from a small subset of participant (5) was also included in the data package, but was ultimately excluded from the analysis as irrelevant to this study's scope.

189 individual .CSV (text) files, one per Driver attempting (but not necessarily completing) the study provided all logged timestamped GPS locations. On-board loggers were programmed to record vehicle position every 30 seconds while in operation, starting with car engine on, and ending with car engine off (Figure 2). Points logged between a single engine ON/OFF period were grouped into a single trip, and logged separately as a line in the Trips spreadsheet, which also contained trip-averaged values of basic variables of interest (such km-travelled, duration of travel, accelerations and deceleration events (hard and total), and fuel economy/consumption). The Trips summary spreadsheet and associated .CSV files were the main sources of data used in this analysis, and initially included more than 2.5 million logged points combining into 60,400 trips.

Neither information on how the various fields in the spreadsheets and text files related to one another, nor actual metadata on the context and intent behind the collection of individual fields was provided, for example as part of the original study design documents. Information such as this is critical in any subsequent analysis; had it been produced prior to the start of the study, it would have allowed for more efficient data collection and recording, and a lighter data transformation load. The first step to the analysis was therefore an attempt to derive both through exploratory statistical analysis and visualization.

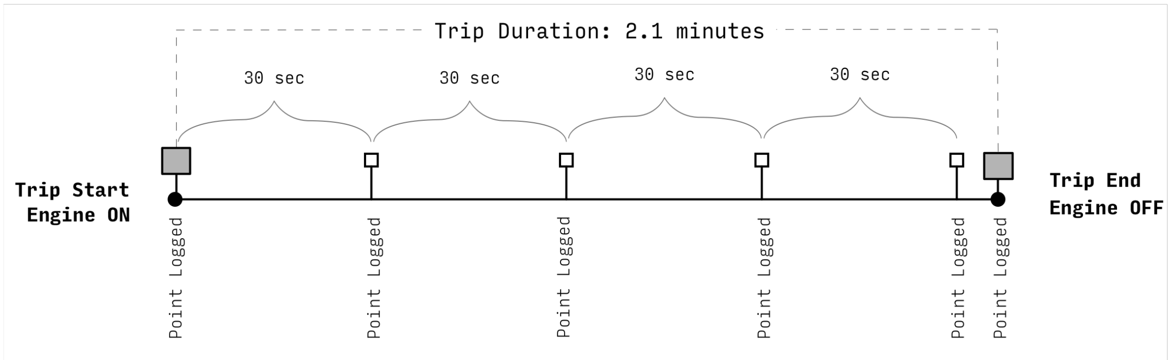

Figure 2: *Logging GPS locations and summarizing trips.*

Data Transformations & Storage

Once the primary variable of interest and relevance to the study had been identified through the preliminary exploratory analysis, the data was scrubbed, reclassified, and filtered to produce the subsets used in the analysis. In its original form, the provided data was not in a format suitable for extensive analysis. Extensive clean-up and transformations included removing duplicate fields and NULL/empty rows, validating classifiers (e.g. use of number 3 rather than word "three" used by some survey fields), as well as replacing duplicate text-based entries (e.g the education level: ["College, Apprenticeship or Trades Certificate or Diploma or University certificate or diploma below bachelor level"] with the numerical index [2]) and connecting said field to a secondary definitions table. The transformed data was subsequently filtered, exported as text field (in .CSV), transformed (to .SQL) and input into a local MySQL database specifically designed and created to store it. The data processing steps described above are summarized by the flowchart in Figure 3. Figure 4 is the Entity-Relationship diagram drawn up for the purposes of designing the database, and the Table diagram in Figure 5 documents its structure. The creation of the MySQL database was an important milestone, as from this point onwards, the data which was of higher overall quality than the original, would be accessible by various analysis and visualization tools, and could be stored and distributed centrally for future projects, if necessary.

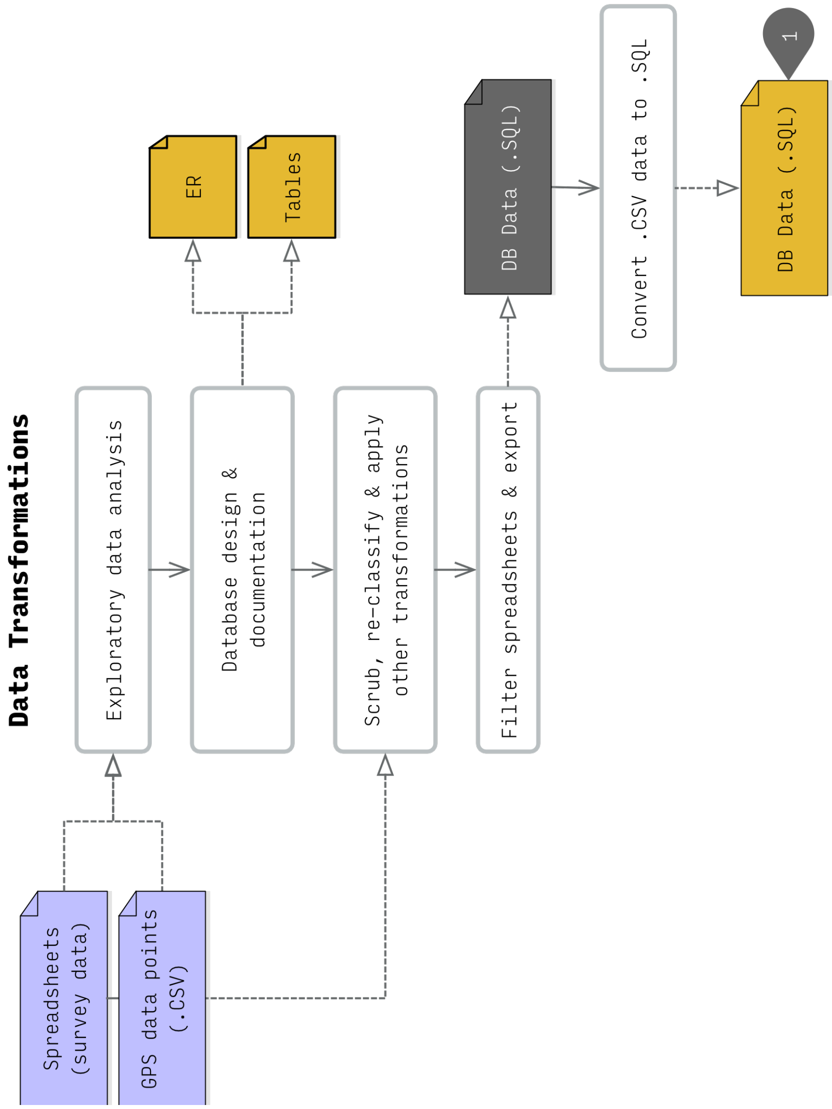

Figure 3: Flowchart illustration of the Data Transformations stage of the approach.

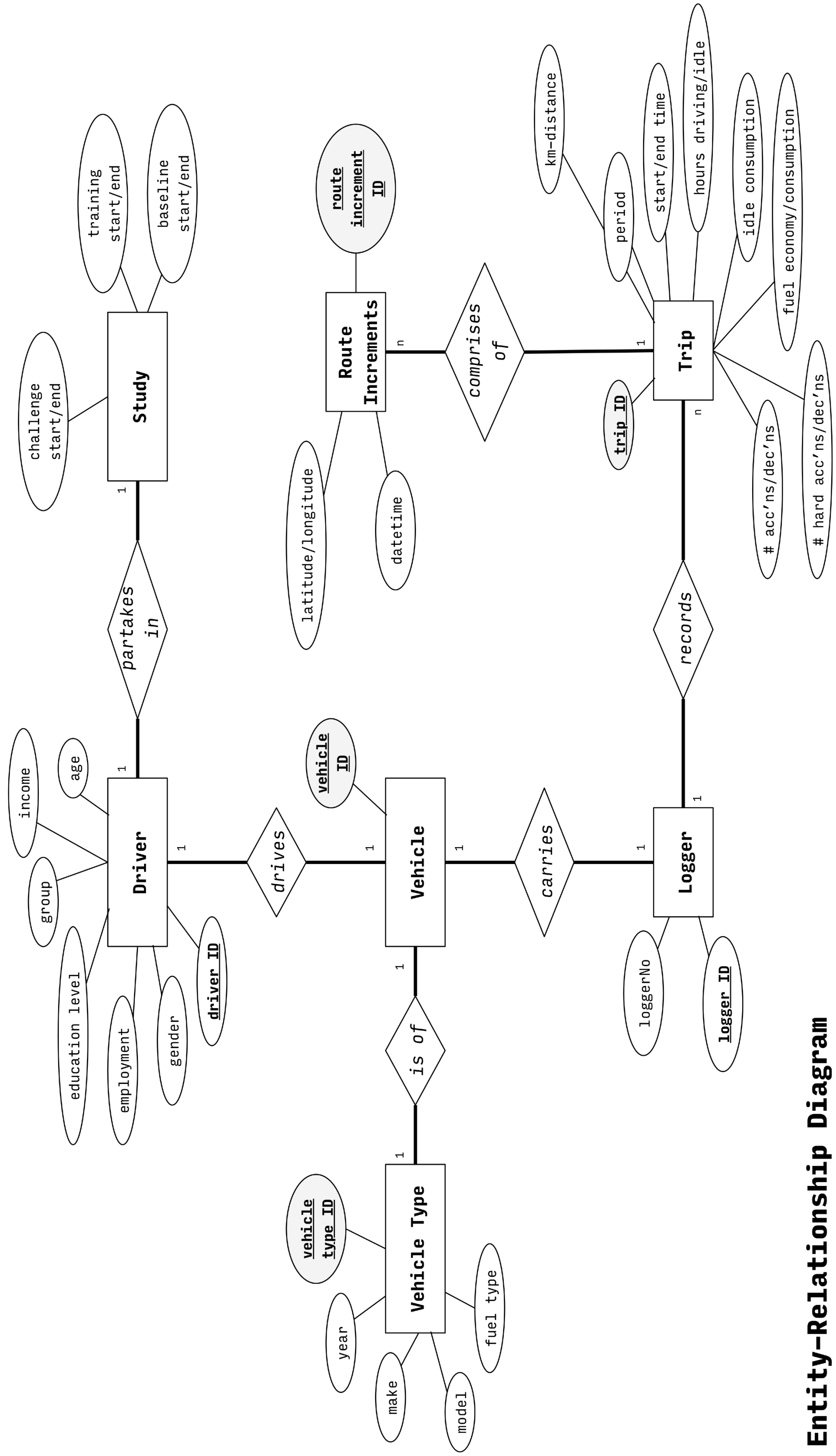

Figure 4: *Entity-Relationship diagram for the data used in this study.*

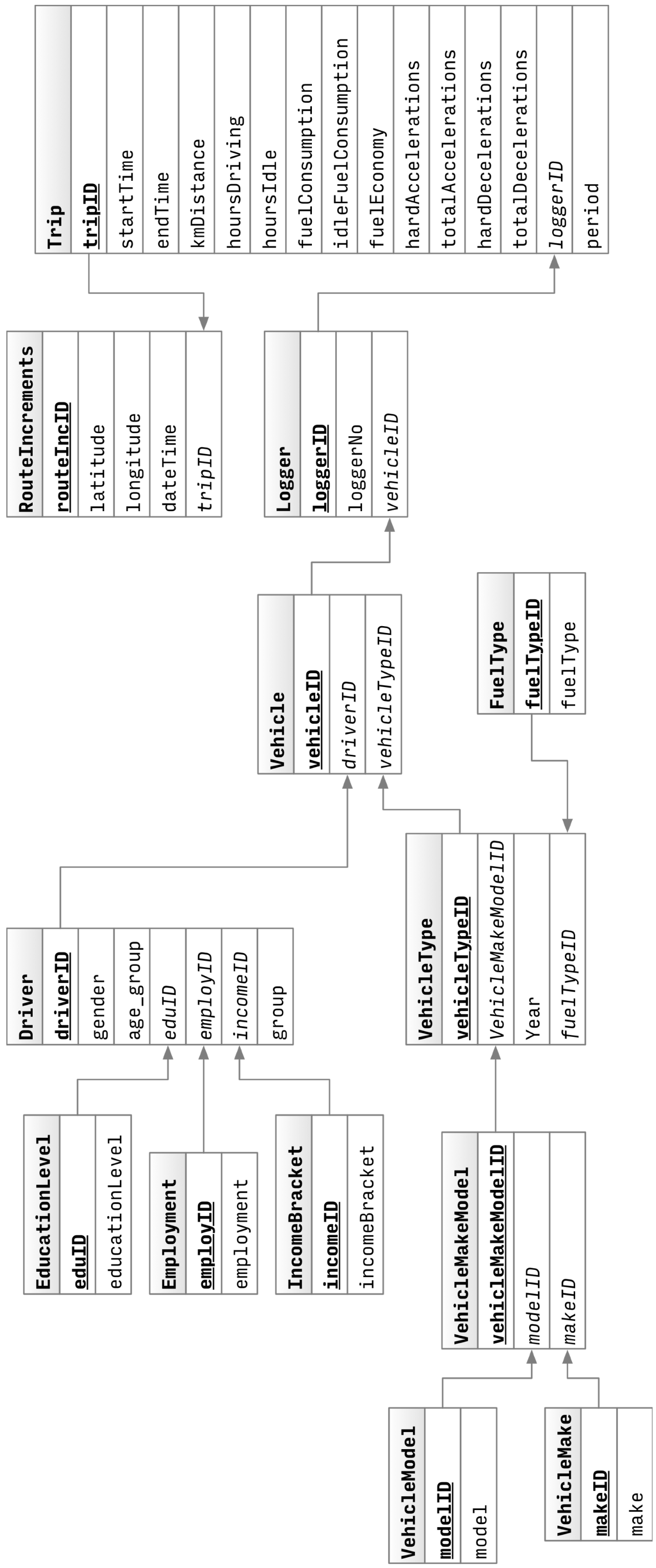

Figure 5: *Table diagram for DB holding the data used in this study.*

Prior to mapping the data two transformations had to be applied to the data. The first related to the fact that although the provided GPS data were linked to the individual logger that recorded them (by use of a unique, non-numeric, logger number [:loggerNo]), there was no link to the actual trip they belonged to, and for which summary values of the key metrics were calculated for (automatically, by the on-board loggers). The only other fields were the timestamp and the Lat/Lon position. A connection for the individual trip each GPS entry was a part of was critically important, as it would impart necessary contextual information to the mapped points, allowing for subgrouping, layering, geographical analysis and visualization.

An automation script was written to make this connection, by comparing each point's timestamp with the start and end timestamps of logged trips with matching loggerNo's. To do this, both trip and route increment points were assigned a unique numerical ID, which also served as the PRIMARY key in their respective tables. If the point timestamp fell within a start and end date time of a particular trip, the corresponding tripID from the Trips table was added to the RouteIncrements table, as a FOREIGN key linking the two (Figure 6).

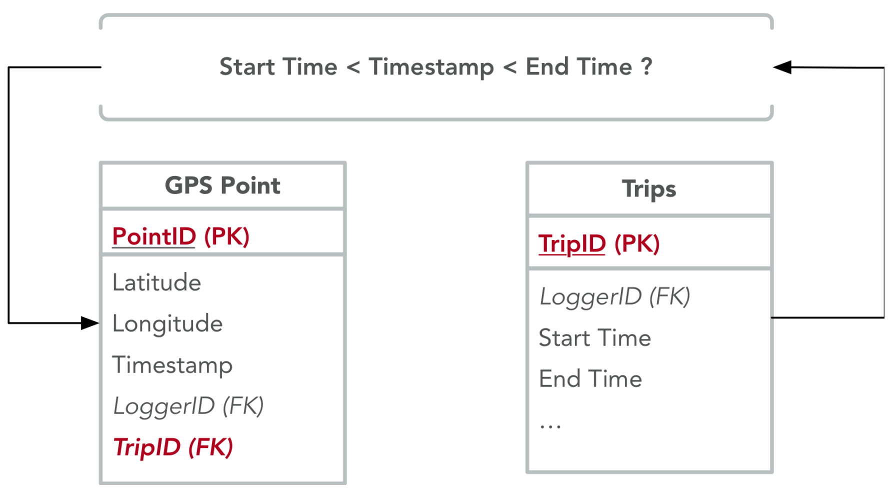

Figure 6: *Connecting the path/route points to Trips.*

The second transformation was to determine the period each trip belonged to (i.e. one of Baseline, Training, or Challenge) — a critical piece of information in the analysis of pre- and post-training changes. This step was also completed through scripting, by comparing the start and end dates of each trip with the corresponding Training start and end dates in the Participation spreadsheet data:

Determining trip period:

```
IF ([Trip].[End Date] < [Driver].[Training Start Date])
       THEN Period = "BASELINE"
ELSE IF ([Trip].[Start Date] > [Driver].[Training End Date])
       THEN Period = "CHALLENGE"
ELSE Period = "TRAINING"
```

The corresponding period was then entered as a new field named period in the Trips table. The flow chart in Figure 6 illustrates this process.

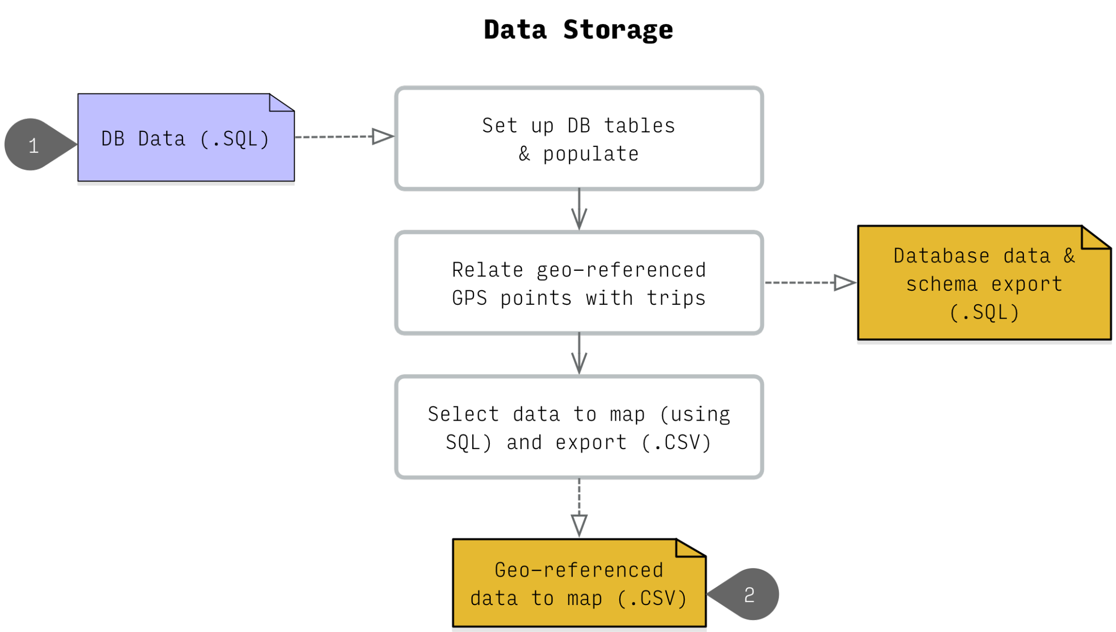

Figure 7: *Flowchart illustration the Data-Storage & Linking stage of the approach.*

Methods/Tools

The specific tools used in this study are listed in Table 1. All of the work was conducted on an iMac (27-inch, late 2012), running OSX El Capitan (10.11.6), with the following specs: 3.2 GHz Intel Core i5, 32 GB (1600 MHz) DDR3, NVIDIA GeForce GTX 675MX 1024 built-in graphics), 1TB SATA SSD. Any MS-Windows-exclusive applications (namely ESRI ArcGIS, not available on OSX) were run through Parallels Desktop 11 for Mac (v. 11.1.2).

Table 1: *List of tools and their use in this project.*

Application/Tool	Description of Use	Cost	OS
R (Studio)	*Cleanup/scrubbing, transformations, preliminary graphs, statistics.*	Free	
MS Excel	*Cleanup/scrubbing, transformations, preliminary graphs, statistics.*	$$	
Perl	*Text processing – big data files.*	Free	
Command Line/ Terminal	*Text processing – big data files.*	Free	
Tableau Desktop Pro	*Visual analytics (exploratory + post-processing analysis), mapping, time-lapses. Used Student Licence.*	$$$	
ESRI ArcGIS	*Mapping, geographical analysis (e.g. near tables, distance-based filtering, identification of major thoroughfares). Used trial version.*	$$$	
ArcPy (*with ESRI ArcGIS*)	*Task automation. Used within ESRI ArcGIS (trial version).*	Incl.	
MAMP	*Apache, PHP & MySQL servers, run as apps.*	Free	
MySQL	*DB setup & data storage/retrieval, querying, subsetting, reclassification. Included in MAMP.*	Free	
PhpMyAdmin	*DB interface utility for MySQL. Included in MAMP.*	Free	
BigDump.php	*Bulk/large file size data uploads (SQL or CSV) to phpMyAdmin.*	Free	
Adobe Photoshop	*Finalizing maps, image processing (bitmap graphics). Used trial version.*	$$$	
Adobe Illustrator	*Finalizing maps, image processing (vector graphics). Used trial version.*	$$$	
AppleScript	*Task/code automation. Part of OSX.*	Free	
Automator	*Task/code automation. Merging CSVs. Part of OSX.*	Free	

4 Analysis & Results

The scrubbed and transformed data was subsequently examined in different ways, namely: statistically (using MS Excel and R) to describe it; geographically (using ArcGIS) to visualize the physical extent of all recorded paths, and to identify any geographical outliers; and visually (using Tableau) visually identify shifts in emerging pre- and post-training driving patterns (Figure 8).

Geographical overview

Mapping revealed issues not previously discernible by viewing the data within as rows cells of a spreadsheet or database table. For example, the very first observation made on the mapped trips was that a number of points were shown to have taken place in Toronto, ON. A more detailed inspection indicated that these points belonged to 2 separate loggers and were actually logged along physical street boundaries, indicating they were potential left-overs from previous usage, rather than calibration errors. These points were ultimately filtered out.

An implication of the Smart Drive Challenge project conducted by Metro Vancouver was that the study scope concerned car use and driving emissions within the MVRD, and within the context of regional (rather than provincial/long-distance) travelling. Study participants were not constrained with respect to the duration, physical extent, or purpose of travel, with many drivers traveling past the administrative boundaries of the MVRD. These extra-regional trips were also logged outside BC (e.g. Calgary, AB) and in the USA (Seattle, WA and Portland, OR). These trips were much longer on average compared to those confined within the region, and they had to be excluded from the analysis so as to not obscure small but not insignificant changes in driving patterns for drivers staying close to their base.

To this end, an area limit that comprised of the MVRD + Abbotsford was defined in ArcGIS and used to intersect all data points (Figure 9). Abbotsford was added to the area limit because of the number of trips that were logged within its boundaries.

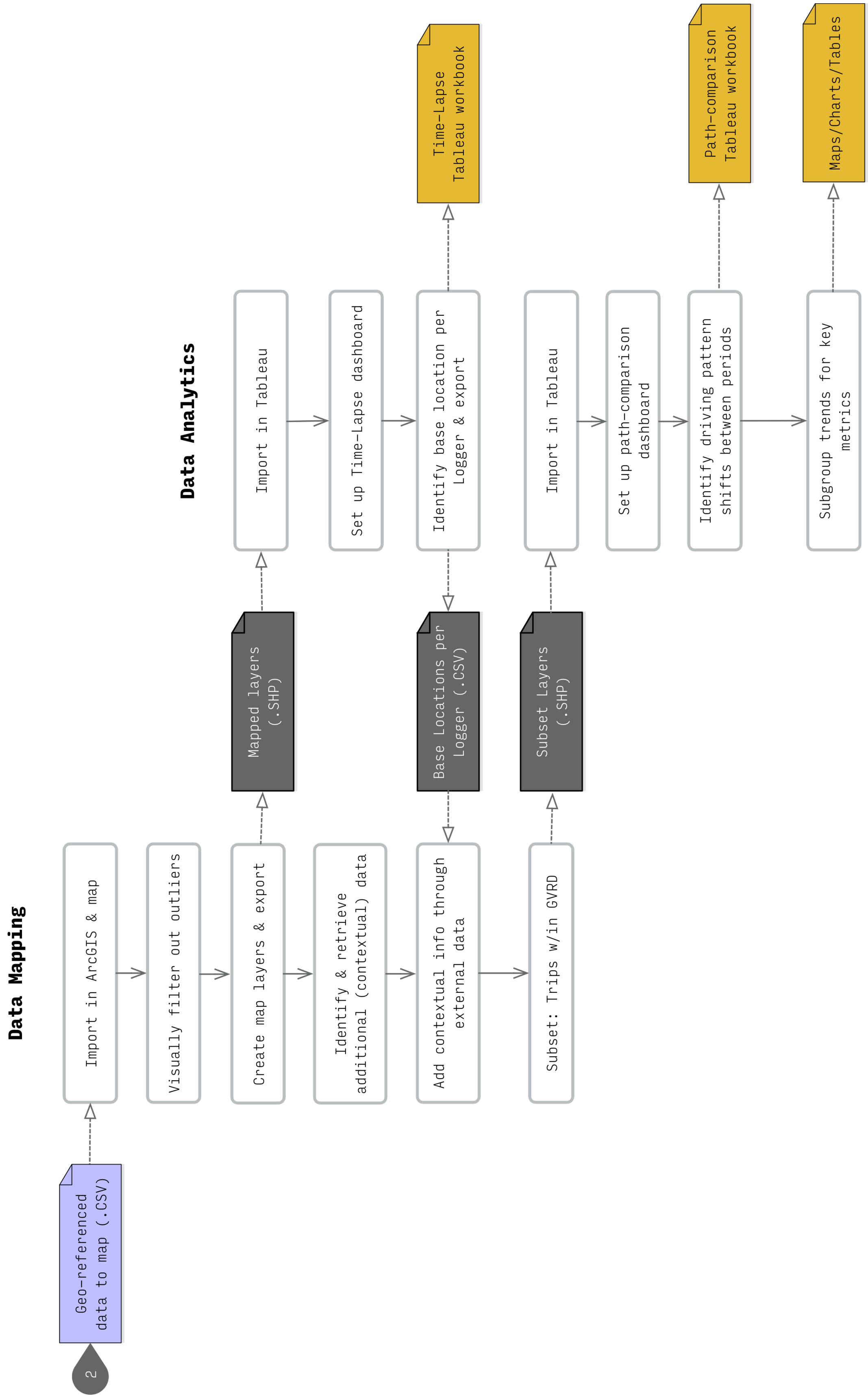

Figure 8: *The Mapping and Data Analytics stages of this study.*

A list of unique tripIDs was extracted from the intersected data, and used to separate trips into 2 groups, one for the region, and another for extended-travel. A total of 1572 trips fell outside the cut-off boundary and were excluded, even if they had originated within it. Figure 10 maps the trips that were enclosed by the cut-off area in their entirety and were therefore included in the analysis (subset "IN"). All subsequent analysis and results presented below were conducted on this regional ("IN") data subset.

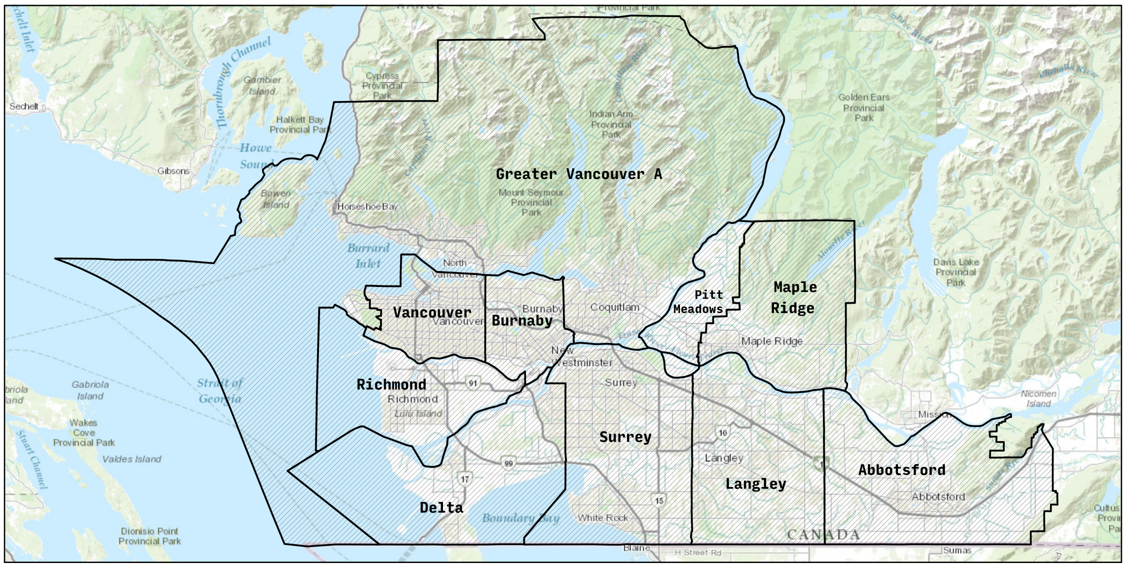

Figure 9: *Exclusion/cut-off area for regional trips.*

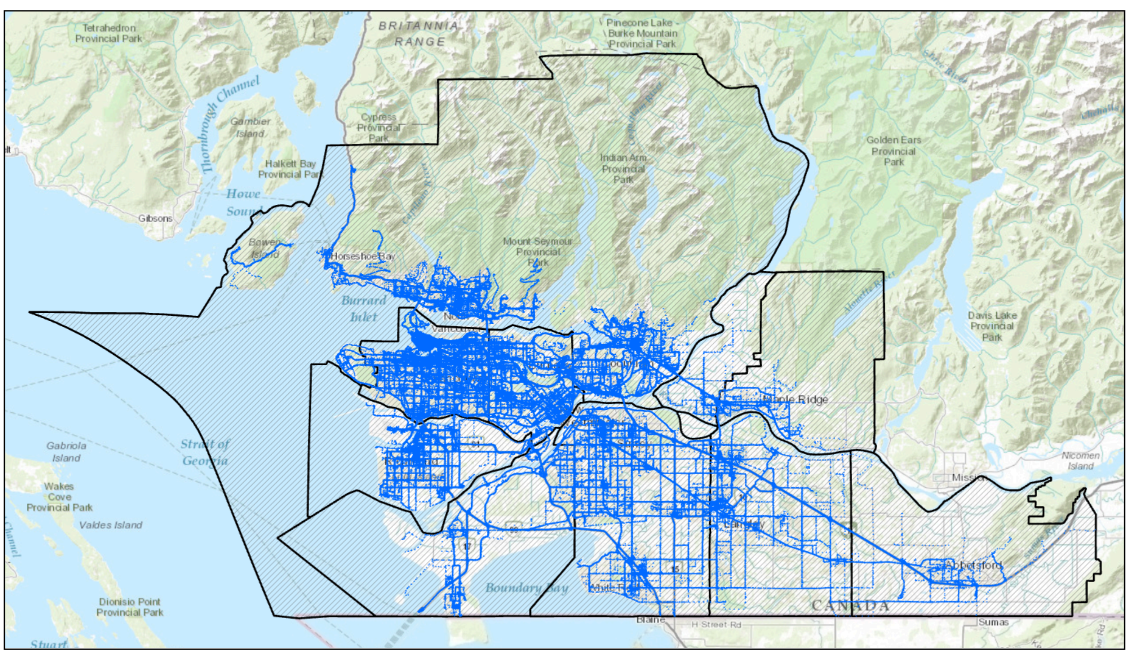

Figure 10: *Trips enclosed by the MVRD-focused exclusion boundary that were included in this analysis.*

Analysis & Mapping

From the available fields in the database, 25 were of particular interest and used to define a set of metrics that could be used to address primary question: *"Can we drive less?"*. They were:

(1) Average number of Trips per day

(2)-(4) Total hours driving, per-trip-average hours driving, and per-period-average hours driving;

(5)-(7) Total hours idling, per-trip-average hours idling, and per-period-average hours idling;

(8)-(10) Total, per-trip-average trip duration, and per-period-average trip duration;

(11)-(13) Total, per-trip-average, and per-period-average km-travelled;

(14)-(15) Average and max trip speed;

(16)-(17) Average fuel consumption (in [L]), and average fuel economy (in [L/100km]);

(18)-(25) Total accelerations, total hard accelerations, total decelerations, total hard decelerations, per-period-average hard acceleration, per-period-average hard deceleration, hard/total acceleration ratio, and hard/total deceleration ratio.

The analysis comprised of calculating the changes in these 25 metrics, between the Baseline and Challenge periods, for the filtered (IN) subset. Due to the excluded OUT subset trips, the duration (in calendar days) of Baseline and Challenge periods covered by the included data was not the same, neither between periods, nor among participants (Figure 11) — it had to therefore be accounted for when calculating metric averages.

The calculated pre- and post-training changes in these metrics are presented in Figure 12. The resultant values had either a positive or negative numerical value, depending on whether a particular metric increased or decreased between periods, and some negative values had a positive connotation, for example a reduction in average fuel consumption indicated less time on the road. In the bar chart of Figure 12, changes with a positive connotation are marked in green, and changes with a negative connotation are marked in red. Changes in trip duration for which a numerically positive change can be interpreted either way, was marked in blue.

The most significant changes in overall trends were an average reduction in hard accelerations and decelerations (by -36.05% and -31.14% respectively). There also was a (positive) reduction in both the per trip and per period average km-distance travelled (by -4.9% by -10.56% respectively). Although both the total and per trip average hours driving increased (by 5.03% and 1.73% respectively) — not ideal w.r.t. the goal of driving less — the per period average hours driving was actually reduced by -2.81%. In a similar fashion, although the total and per-trip-average idling hours increased (by 6.44 and 2.65% respectively), the per-period-average idling hours dropped sightly by -1.23%.

The correlation matrix of Figure 13 was created in R based on the calculated metrics presented above. The scale in Figure 14 was used to describe the linear relationships between metric pairs based on their calculated correlation coefficient (r) values. The metric pairs and their corresponding r values are summarized in Figure 15. More specifically:

— Trip duration (derived from the total count of route increment points) was indicated to be ***directly positively correlated*** with logged hours driving.

— Trip duration was also indicated to be ***strongly positively correlated*** with logged hours idling and hard decelerations. Km-distance-covered was indicated to be *strongly positively correlated* with logged hours idling, logged hours driving, trip duration, average speed, hard accelerations, and hard decelerations.

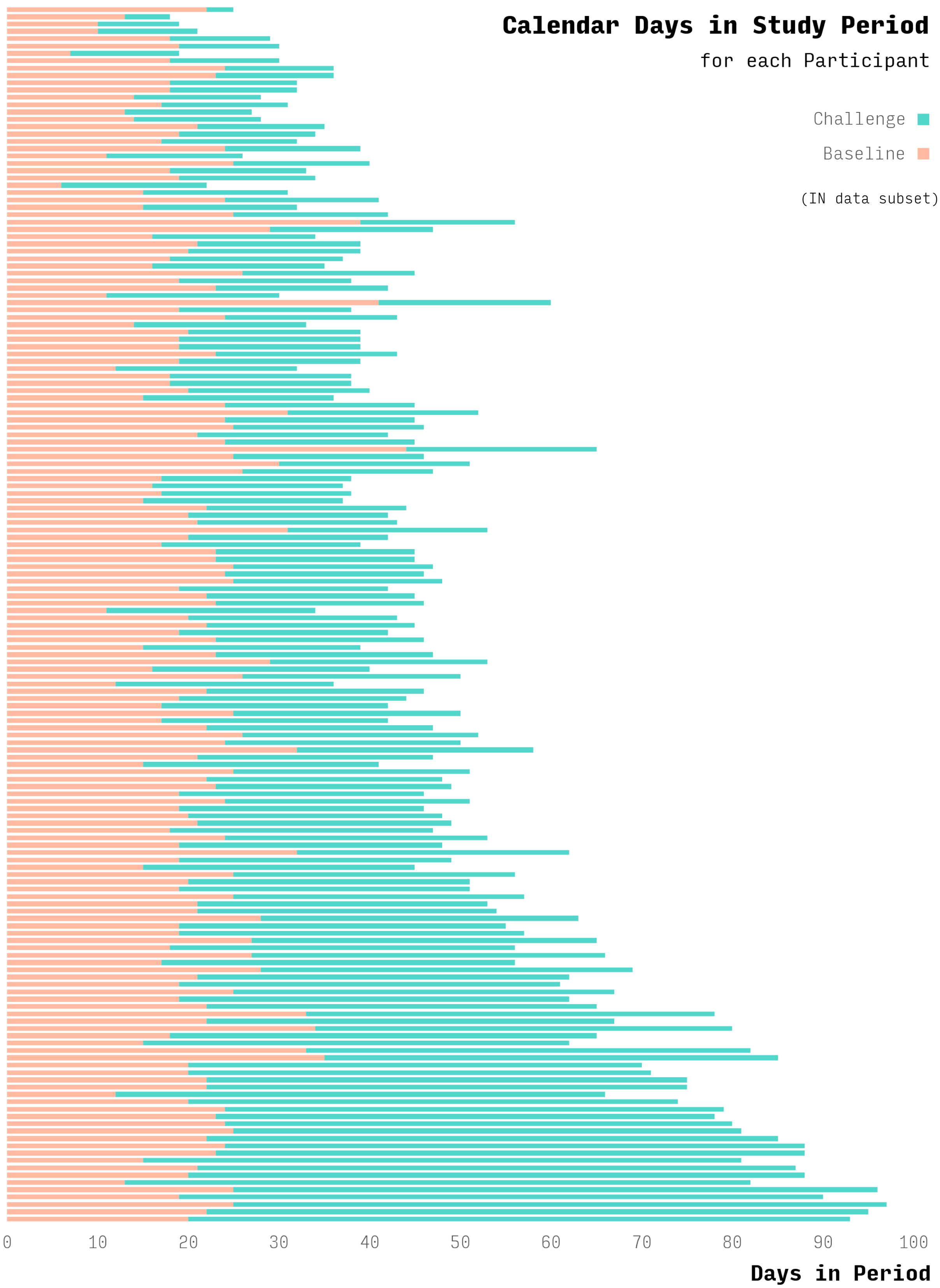

Figure 11: *Calendar days in study period per participant.*

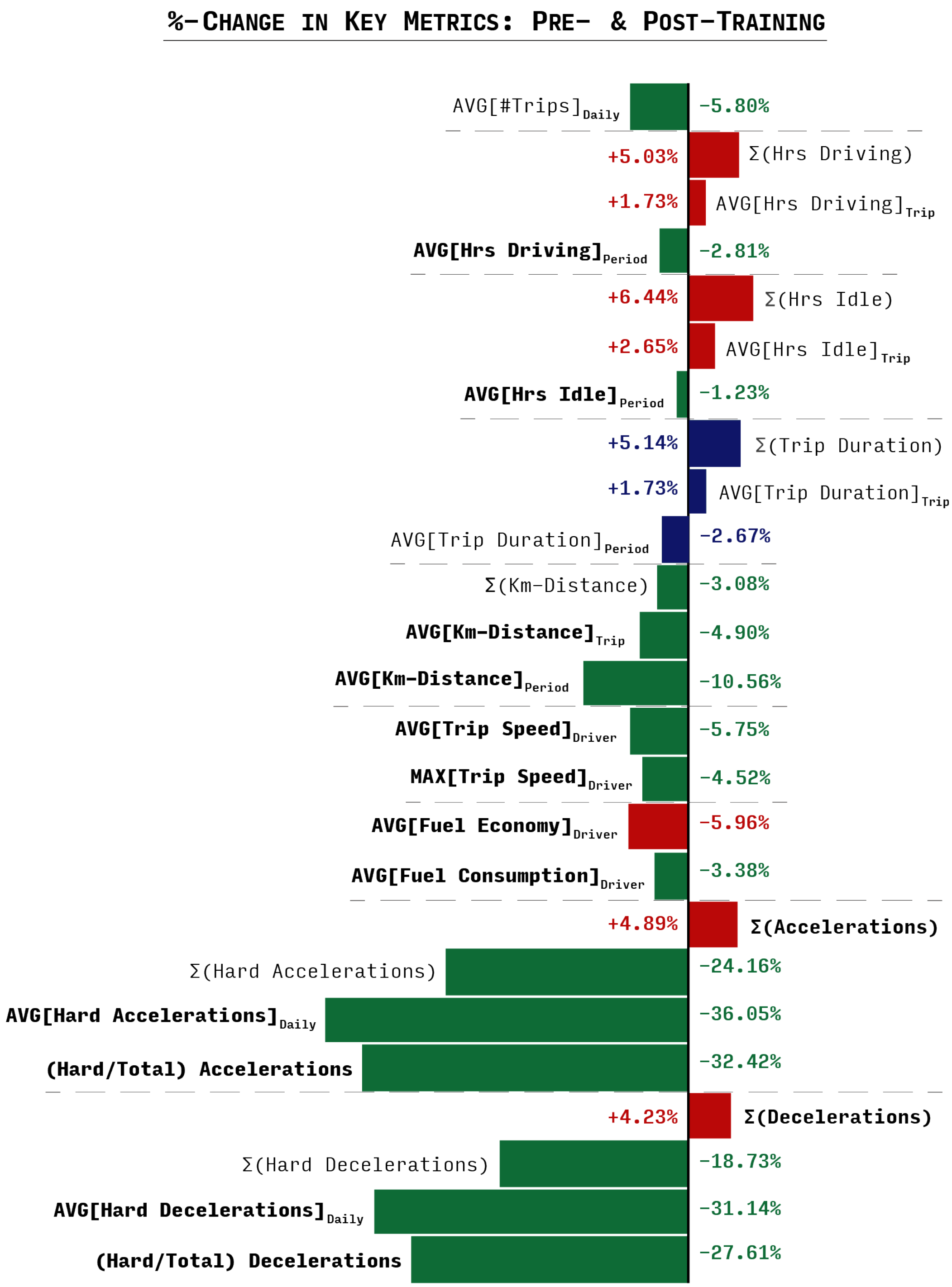

Figure 12: *Pre- and post-Training %-change in 25 key metrics used to assess changes in driving behaviour.*

- Fuel consumption was indicated to be ***strongly positively correlated*** to average speed. Hard decelerations were indicated to be *strongly positively correlated* to hard accelerations, and hours driving. Finally, logged hours idling were indicated to be *strongly positively correlated* to hard accelerations and hard decelerations.

- Fuel consumption was indicated to be ***moderately positively correlated*** to idle fuel consumption, while hard accelerations and decelerations (period averages) were indicated to be *moderately positively correlated* to km-distance and hours idling.

- Hours idling were indicated to be ***weakly positively correlated*** to idle fuel consumption. Average speed was indicated to be *weakly positively correlated* to both km-distance and fuel consumption.

The change (marked as Δ) in period averages for the 4 metrics that are more closely associated to the concept of driving less (i.e. trip-count, km-distance, hours-idling and hours-driving) were calculated for 5 driver attribute variables: Gender, Age, Employment Status, Education Level, and Income Bracket. They are presented in Tables 2-6. Table 7 summarizes the same 4 metrics based on reported area of domicile.

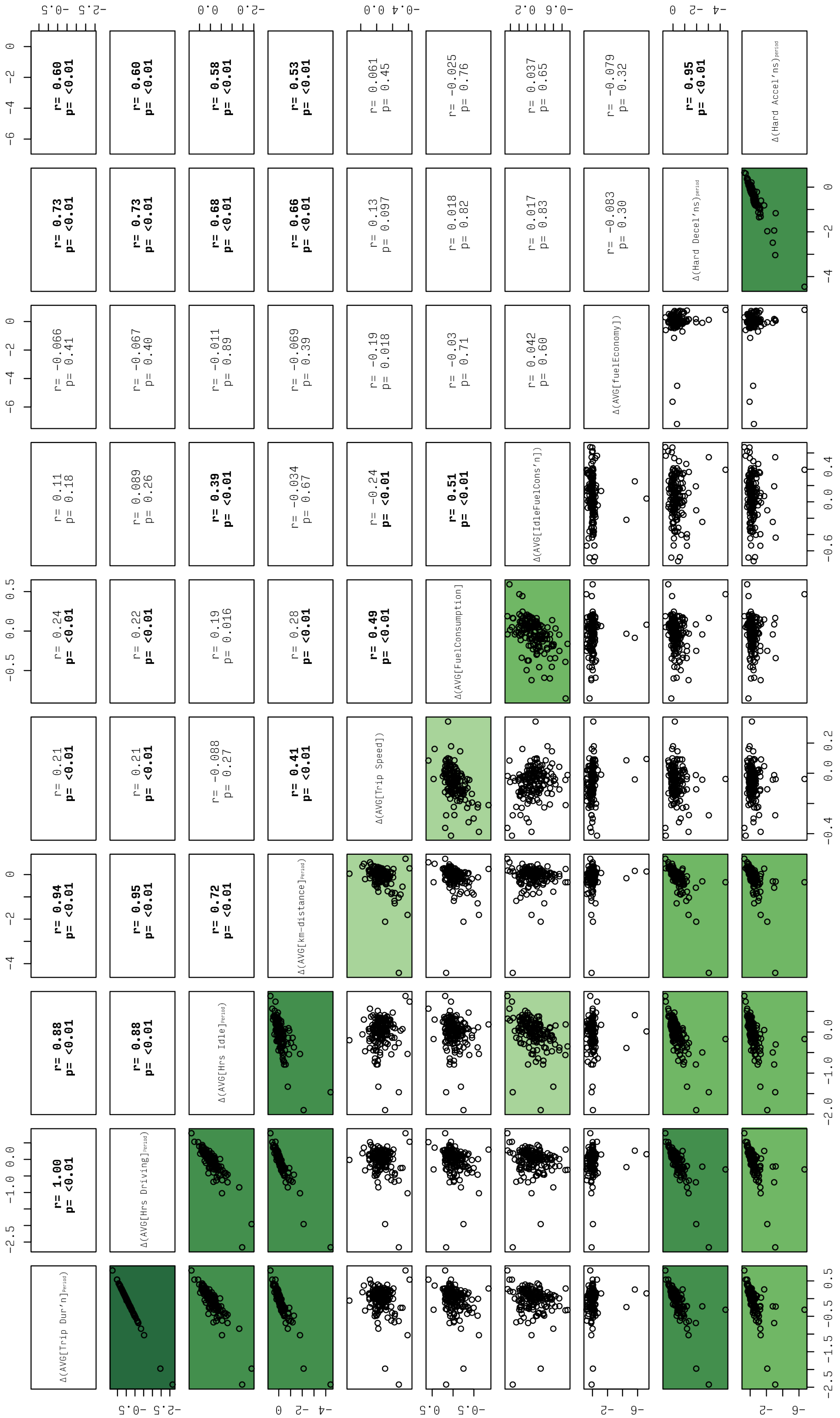

Figure 13: *Correlation matrix.*

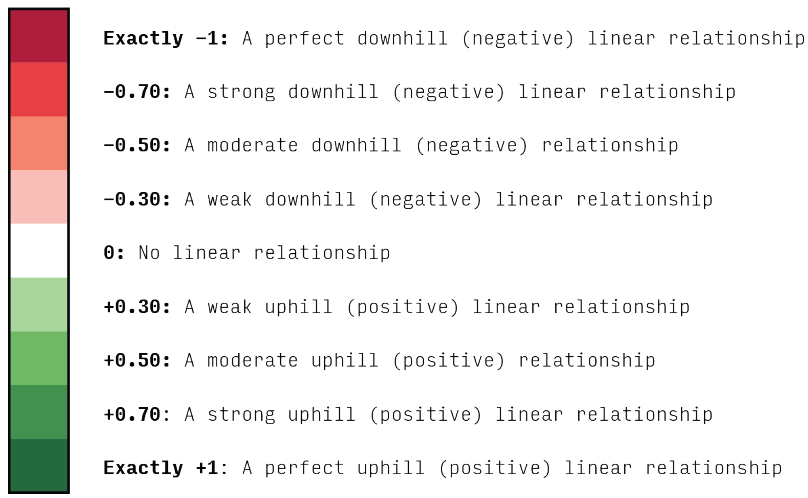

Figure 14: *Interpretation scale used to assess the correlation between attributes, based on the correlation coefficient (r)*[1].

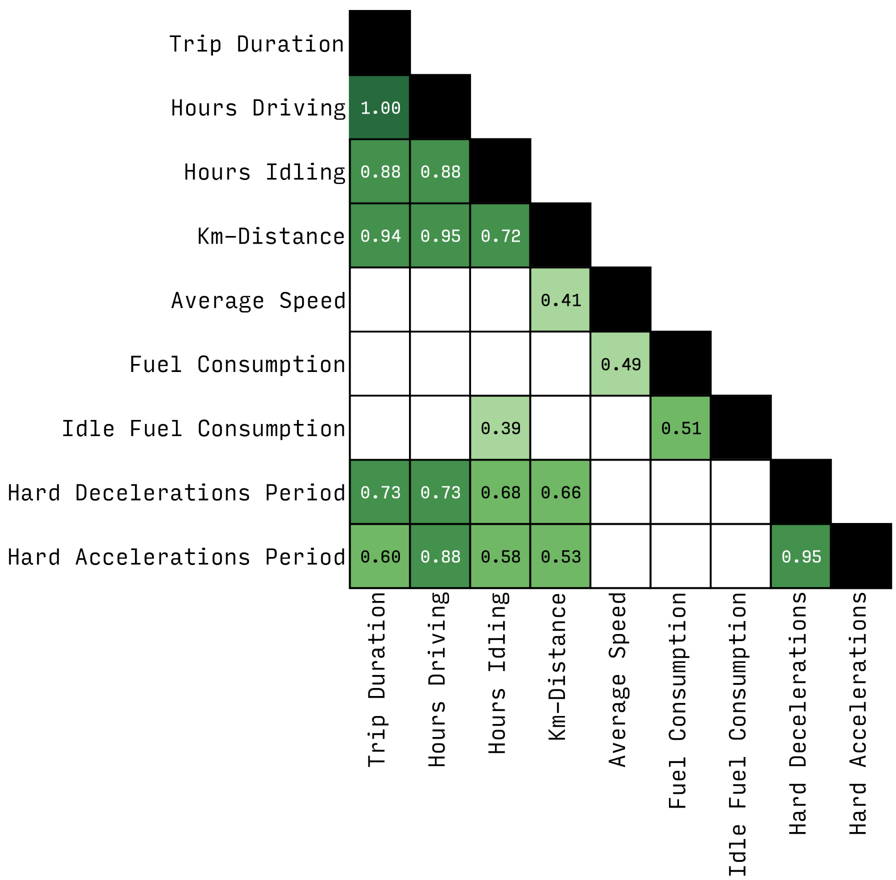

Figure 15: *Correlation values for select variables; cell shading based on interpretation scale of Figure 21.*

[1] Adapted from: Rumsey, Deborah J. 2016. Statistics For Dummies. Hoboken, N.J.: John Wiley & Sons.

Table 2: *Average change in trip count, km-distance, hours idling, and hours driving, by gender.*

Period AVG's:

Gender	Δ[Trip Count]	Δ[Km-Distance]	Δ[Hours Idling]	Δ[Hours Driving]
Female	-3.89%	-5.49%	-0.34%	0.44%
Male	-7.14%	-14.10%	-1.85%	-5.08%

Table 3: *Average change in trip count, km-distance, hours idling, and hours driving, by age bracket.*

Period AVG's:

Δ[Age]	Δ[Trip Count]	Δ[Km-Distance]	Δ[Hours Idling]	Δ[Hours Driving]
20-29	7.05%	4.49%	8.12%	7.77%
30-39	-7.37%	-14.50%	-3.72%	-5.08%
40-49	-22.33%	-26.77%	-7.14%	-14.00%
50-59	-5.48%	-8.54%	-3.17%	-2.97%
60-69	16.93%	7.70%	6.04%	11.17%
70-79	11.46%	14.10%	17.19%	18.76%
Over 80	26.67%	10.93%	30.45%	23.20%

Table 4: *Average change in trip count, km-distance, hours idling, and hours driving, by education level.*

Period AVG's:

Education	Δ[Trip Count]	Δ[Km-Distance]	Δ[Hours Idling]	Δ[Hours Driving]
College Below Bachelor Level	-6.54%	-5.32%	-1.95%	0.63%
Completed High School	-0.76%	-4.32%	3.71%	1.27%
Less than High School	2.68%	10.00%	26.47%	15.18%
University at bachelor Level	-0.98%	-4.26%	2.38%	0.83%
University Above Bachelor Level	-14.35%	-26.71%	-8.85%	-12.89%

Table 5: *Average change in trip count, km-distance, hours idling, and hours driving, by income bracket.*

Period AVG's:

Income	Δ[Trip Count]	Δ[Km-Distance]	Δ[Hours Idling]	Δ[Hours Driving]
Under $20,000	4.81%	7.91%	-6.21%	10.83%
$20,000 to $49,999	-7.46%	-5.48%	-9.48%	-4.23%
$50,000 to $79,999	0.72%	-2.70%	7.53%	5.08%
$80,000 to $99,999	0.22%	7.66%	7.99%	10.09%
$100,000 to $124,999	-7.57%	-26.51%	-14.66%	-15.18%
$125,000 to $149,999	4.09%	-17.49%	0.62%	-4.74%
$150,000 and over	-21.30%	-27.87%	-5.41%	-14.89%
Prefer not to answer	-8.92%	-5.69%	-1.07%	-0.61%

Table 6: *Average change in trip count, km-distance, hours idling, and hours driving, by employment type.*

Period AVG's:

Employment	Δ[Trip Count]	Δ[Km-Distance]	Δ[Hours Idling]	Δ[Hours Driving]
Retired/At-home parent/Not working	-2.96%	0.36%	3.97%	3.64%
Self employed	-6.90%	-8.04%	9.26%	0.68%
Student full-time	18.81%	24.06%	24.66%	26.84%
Student full-time Working full-time	-5.26%	-6.26%	8.57%	2.26%
Student full-time Working part-time	-8.33%	-9.46%	-10.18%	-9.02%
Student part-time Retired/At-home parent/Not working	-26.00%	-30.16%	-49.79%	-21.96%
Student part-time Working full-time	17.86%	19.41%	36.09%	31.22%
Student part-time Working full-time Self employed	-8.41%	-21.33%	20.68%	-10.80%
Student part-time Working part-time	14.85%	-12.01%	6.78%	2.27%
Working full-time	-7.69%	-14.70%	-4.11%	-5.80%
Working full-time Self employed	0.15%	-2.78%	-5.30%	0.21%
Working part-time	-2.34%	1.52%	5.17%	5.43%
Working part-time Self employed Retired/At-home parent/Not working	-14.16%	-34.27%	-35.10%	-31.61%

Table 7: *Average change in trip count, km-distance, hours idling, and hours driving, by region.*

Period AVG's:

Region	Δ[Trip Count]	Δ[Km-Distance]	Δ[Hours Idling]	Δ[Hours Driving]
Anmore	-48.90%	-71.60%	-25.13%	-42.78%
Bowen Island	13.71%	19.05%	14.39%	20.30%
Burnaby	-4.48%	-1.10%	-5.44%	-0.62%
Coquitlam	-19.05%	-27.33%	-10.30%	-15.63%
Delta	-4.27%	8.25%	-16.59%	2.33%
Langley Township	-15.74%	-21.12%	-17.73%	-14.81%
Maple Ridge	9.41%	24.44%	32.68%	29.00%
New Westminster	5.09%	2.35%	14.39%	10.67%
North Vancouver City	-5.17%	5.35%	1.38%	8.46%
North Vancouver District	-8.88%	-9.71%	-2.65%	-6.04%
Pitt Meadows	-4.56%	-14.70%	-9.36%	-8.31%
Port Coquitlam	-8.48%	-14.49%	-10.24%	-8.92%
Port Moody	-10.15%	-3.96%	-24.50%	-9.41%
Richmond	-2.56%	-1.57%	2.90%	2.43%
Surrey	0.45%	-2.00%	8.48%	6.15%
Tsawwassen	0.42%	-6.55%	-3.41%	0.16%
Vancouver	4.87%	-6.32%	7.50%	4.56%
West Vancouver	3.86%	-36.62%	-0.37%	-13.85%
White Rock	21.06%	25.79%	33.21%	28.96%

Time-of-Day & Weekday Pattern Shifts

Figures 16, 17 and 18 show 3 versions of a heat map that was created using total period-averaged counts of Route Increment ID's, in 15-minute intervals throughout the day, for each day of the week, and for the 2 periods of interest (Baseline and Challenge). The intent was to see if any pattern shifts in driving through the workday would emerge, which may be of particular interest to public transportation schedulers/planners. The analysis indicated that the most significant shifts occur early in the day and late at night. The 06:45 morning density highs seem to shift about 45 minutes later in the morning time-wise, and from Friday/Saturday to closer to mid-week (Wednesday) weekday-wise (Figure 16). The late-night shifts also push the 10pm night-time density highs about 45 minutes forward but without a corresponding weekday shift (Figure 18).

Path Overlays

Mapping each driver's GPS locations using a very low opacity (here 10%) dimmed infrequently travelled paths made up of few points into the background and brought frequently-travelled paths made up of multiple overlying points to the foreground. Using different colours for each period, it was a possible to compare major shifts in driving patterns resulting from Training and the "smart" feedback. This was achieved using a dedicated Tableau dashboard that allowed period comparisons for individual drivers — also part of this project's deliverables. Although resource/time constraints limited the extent and level of detail that this visual analysis could ideally go into, it was possible to use the dashboard to get a rough understanding of how driving patterns of individual participants varied between periods. Figure 19 shows 4 examples of path comparisons where Challenge paths were more extensive than their corresponding Baseline paths. In the examples of Figure 20, the opposite it true. Expansion after training could indicate that perhaps drivers are grouping more trips together and accomplishing more tasks with every excursion, which would eventually reduce both their fuel consumption and reduce their overall driving hours/km-distance travelled. More detailed analysis on this part of the analysis is recommended.

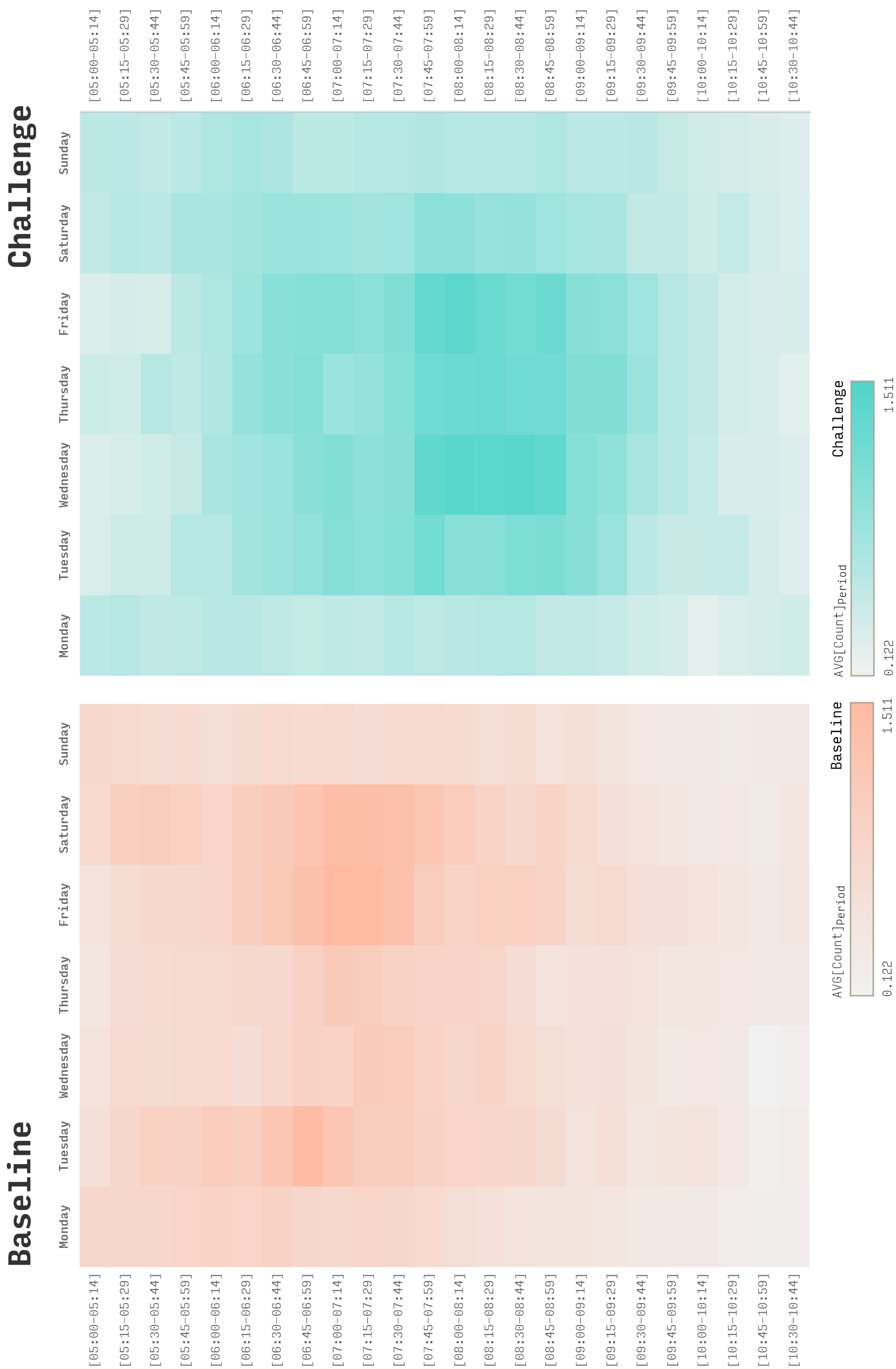

Figure 16: *Shift in patterns through the weekday, Count(Trip Points), [05:00 - 11:00].*

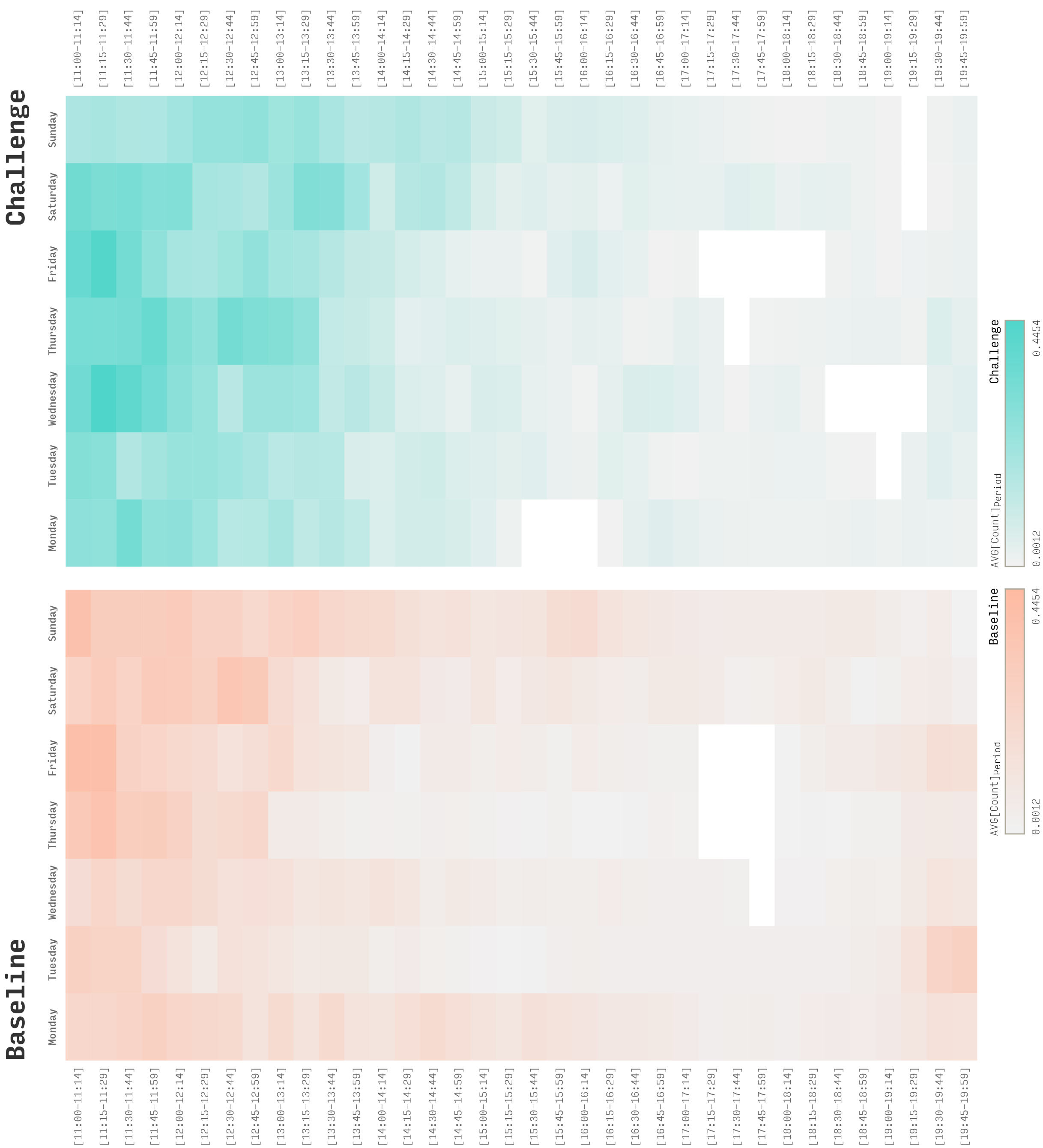

Figure 17: *Shift in patterns through the weekday, Count(Trip Points), [11:00 - 20:00].*

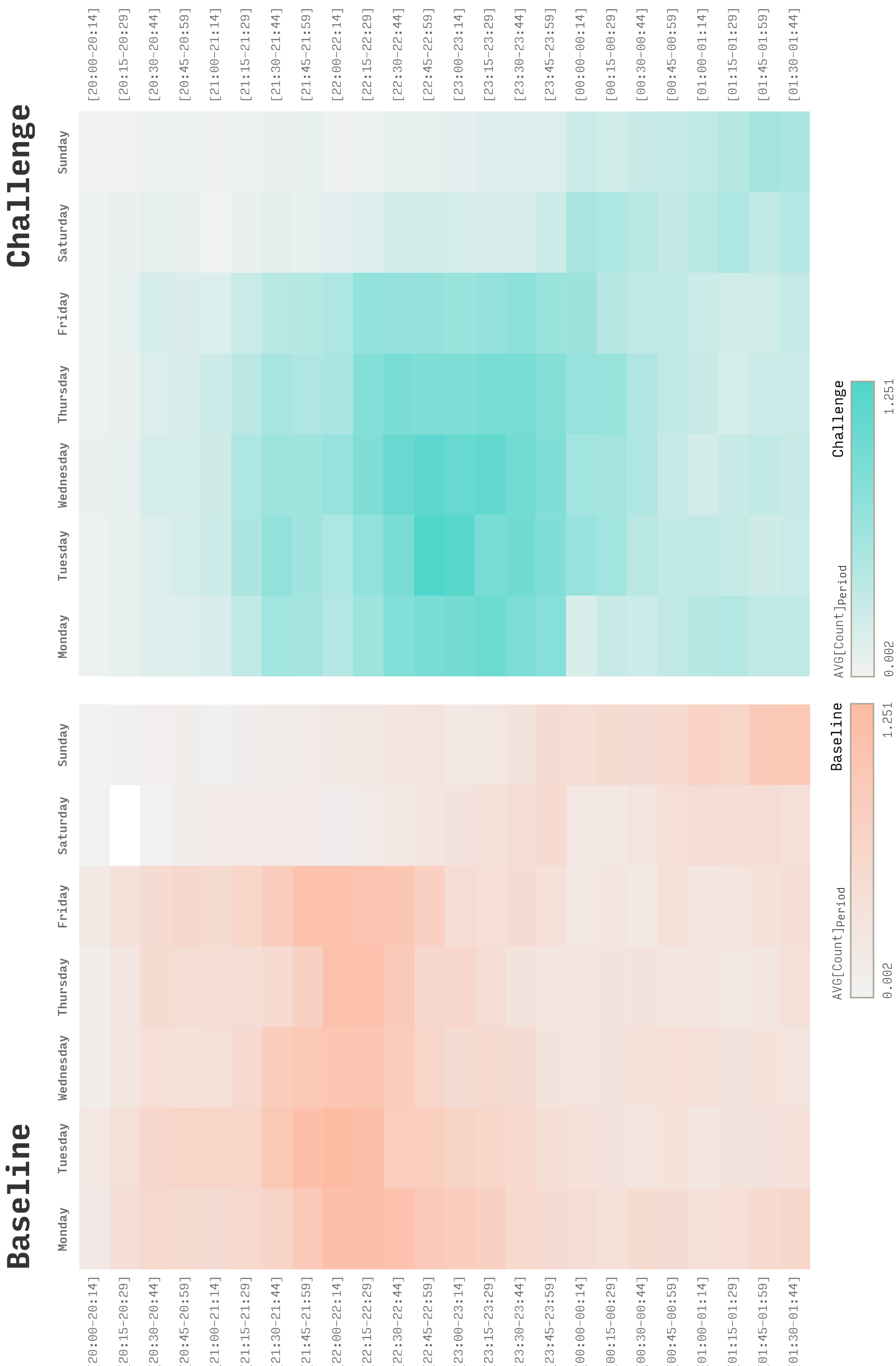

Figure 18: *Shift in patterns through the weekday, Count(Trip Points), [20:00 - 02:00].*

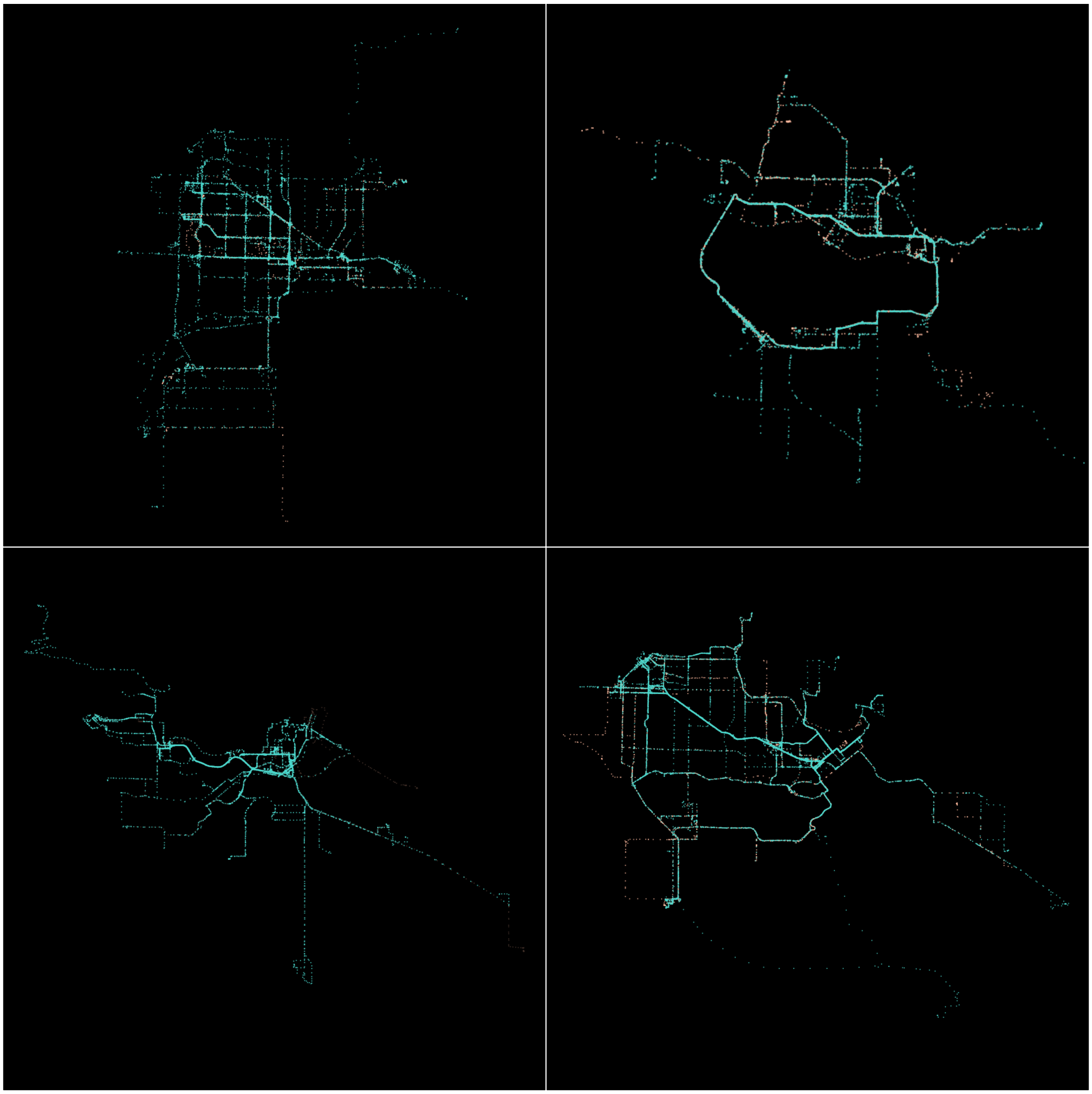

Figure 19: *Example path comparison, with Challenge paths more extensive than the corresponding Baseline pattern.*

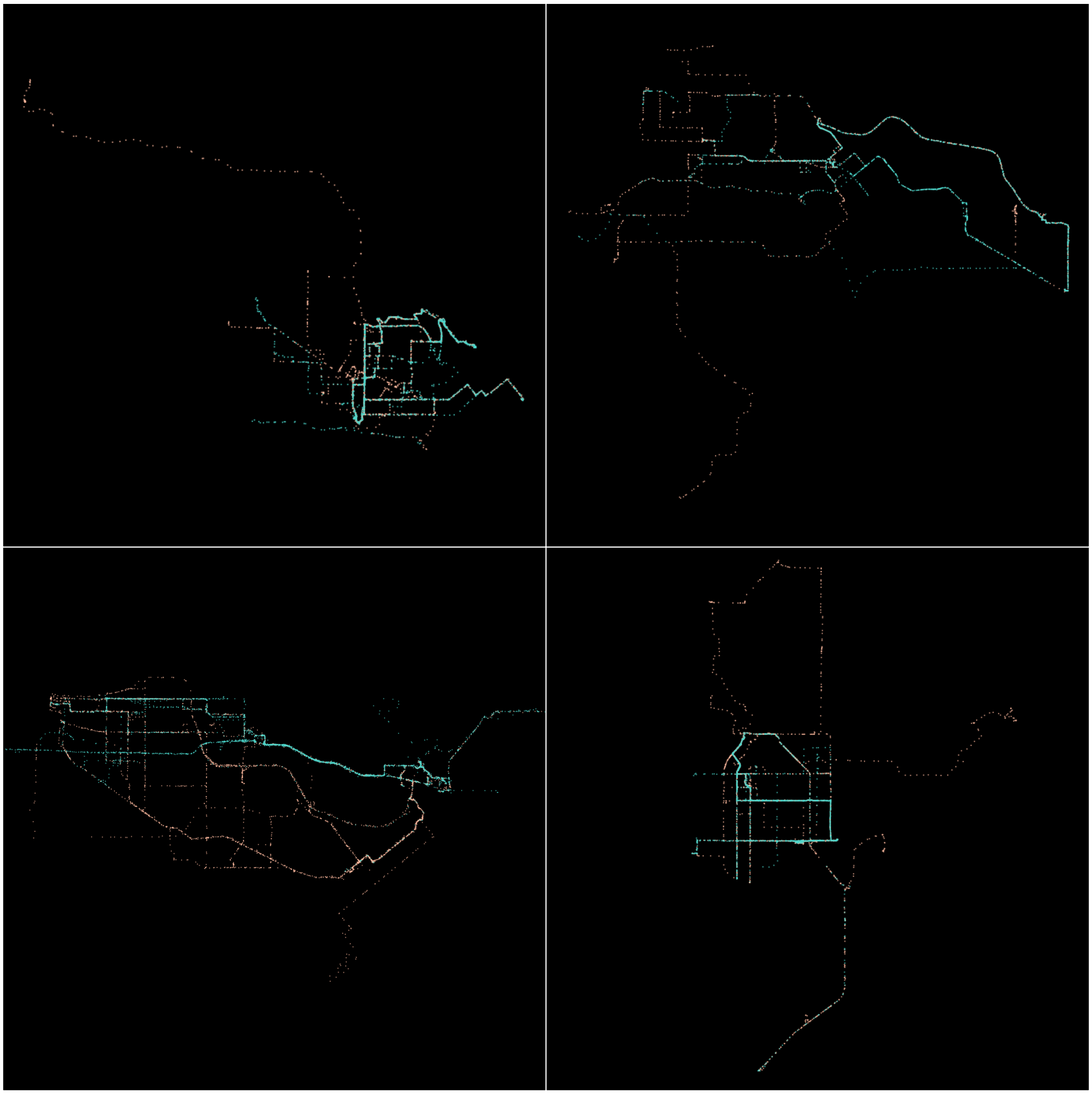

Figure 20: *Example path comparison, with Baseline paths more extensive than the corresponding Challenge pattern.*

5 Conclusion

Summary of Findings

Positive pre- and post-training change in driving habits (as described by the select set of 25 attribute metrics used in the analysis) has been observed, particularly for metrics summarized over the entire calendar-day-duration of each period, for in-city trips whose paths were did not cross the MVRD-focused exclusion boundary of Figure 17. For this particular case, it could be posited that "*we can indeed drive less*".

The most pronounced change in driving behaviour was a reduction in the daily average hard accelerations and decelerations (by -36.05% and -31.14% respectively), as well as in the hard-to-total ratios of accelerations and decelerations (by -31.42% and -27.61% respectively). The overall values of period-averaged km-distance travelled (-4.9%), average trip speed (-5.75%), and average number of completed trips (-5.8%) also dropped. With regards to sub-groupings:

— [Men] achieved a higher reduction in average km-travelled than [women] (-14.10% versus -5.49%).

— Those with [university degrees > Bachelor's level] had the highest km-travelled reduction (-26.71%).

— Age-wise, the age group that responded most positively to the training+feedback was the [40-49] age bracket (with reductions up to -16.77% in all 4 variables (i.e. trip count, km-distance, hrs driving, and hrs idling. The [30-39] and [50-59] age brackets also response positively to the study. The worst response was from the [Over 80] age bracket, with considerable increase in all 4 attributes and most notably hours idling. This may indicate a need for more careful, extensive or customized training for more elderly drivers.

— In terms of income bracket, the [$150,000 and over] bracket responded most positively to the training+feedback, reducing both km-travelled and trip count (by -27.87% an -21.30% respectively).

— [Part time students/At home parents/Not working] was the employment group most positively affected by the training+feedback, with a reduction in hours idling of almost 50%. [Students full-time working part-time], and [working part-time self employed retired/at-home parent/not working] also positively responded to the training. On the other hand, [part-time students working full-time] had the most negative reaction to the training, with a +36% increase in hours idling and a +31% increase in hours driving. [Full-time students] were also negatively affected (up to +28.84% for hours driving).

— Lastly, driver's residing in [Anmore] had the overall most positive response to the training, with a reduction across all 4 variables, most notably km-distance at -71.6%.

Recommendations for Further Work

A number of tasks remain that can potentially be the subject of further research:

— Trips were defined automatically by loggers, with a single Trip corresponding to a single Engine ON-OFF cycle. Changes in the behaviour of drivers are a direct consequence of changes in decision-making, which may be better conceptualized by defining Trips as paths that begin and end in the same physical location, more specifically the driver's base/home. In this way, the lengths of Trip would also become more meaningful in the analysis.

— It may also be of benefit to apply the analysis presented herein to data collected as part of similar studies conducted in regions other than the MVRD (e.g. Toronto or Victoria). Incorporating and/or comparing results from multiple studies could further support this project's conclusions.

— The trips in this study were redefined by uniting any trips that were less than 100 seconds apart. The choice of the 100-second threshold was somewhat arbitrary and more of a practical compromise to deal with a large number of very short trips logged in the data (see histogram in Figure 21). Ideally, had time permitted, it would have been more preferable to redefine the Trips of a sample subset of logged entries manually, and use the redefined trips to assign a more educated guess to the minimum interval threshold for trip redefinition. Such an interval value could also be used in the analysis and processing of data collected by similar studies in the future.

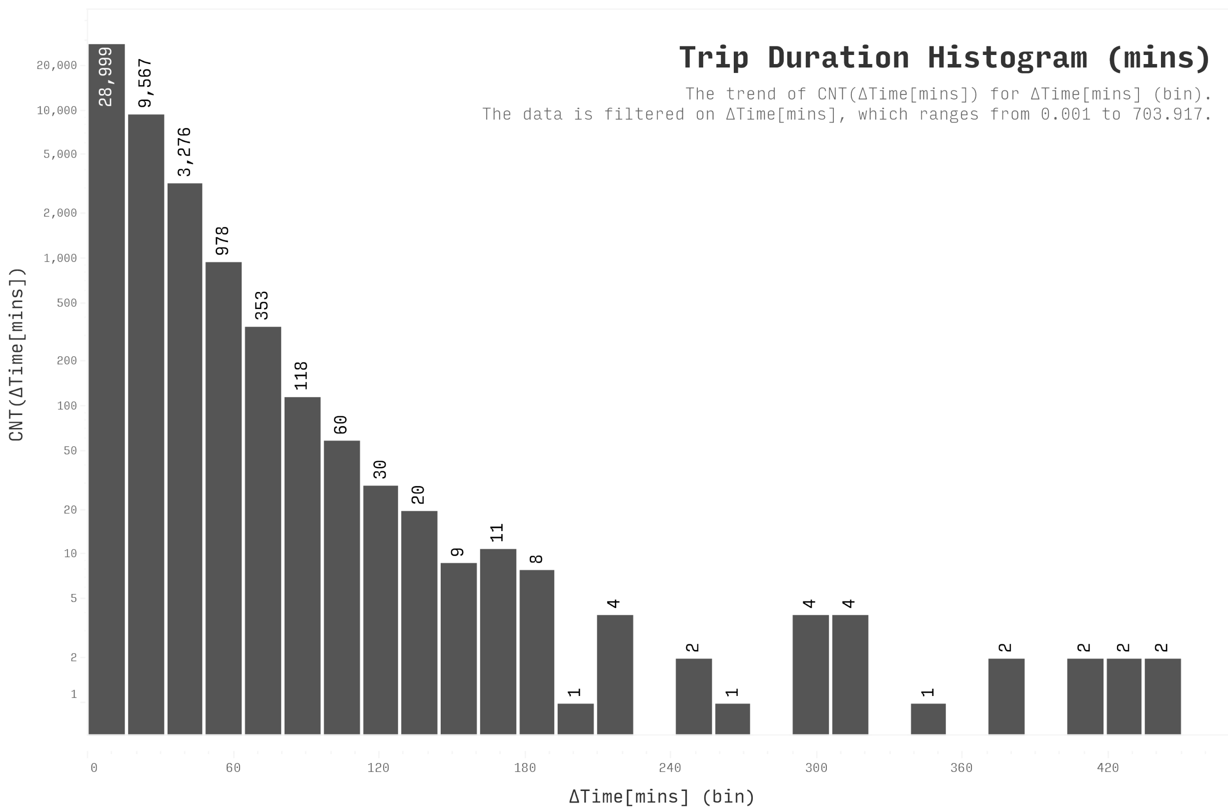

Figure 21: *Histogram of Trip durations, in seconds.*

— The most workload-heady stage of this project was the transformation of the collection of data files into a unified, dynamic, relational database/dataset. Part of the work could actually be done at the collection stage; one serious hurdle this project presented was the fact the individual GPS locations were provided without reference to which Trips they belonged to, which was a critical piece of information that required extensive automation scripting to overcome. The loggers themselves can be programmed to a) assign a unique ID to each Trip they create and store (and for which they calculate a series of summary metrics), and use this TripID as a foreign key in the GIS tables. In addition, as a re-definition of Trips was deemed a desirable preparatory step to analysis, it would also be useful to store the values used to calculate the various Trip metrics (eg. fuel consumption) and allow for the their recalculation based on the original logged points. This would allow for a more accurate re-calculation of these metric, in a manner consistent which whatever method is used to redefine trips (which was not possible in this case, due to this specific lack in data resolution).

— A number of variables (e.g. the number of children in the household) were not included in the initial data subset used in the analysis, due to resource (primarily time) constraints. These fields (particularly the of the survey data, which require extensive processing) should also be processes and added to the core database.

— In terms of data collection during the study, it would be highly beneficial to develop a mechanism to allow for drivers to add contextual information to their travel, either in the form of a travel log, or ideally, as part of the “smart” feedback tech. For example, if participants were able to visualize their GPS data, they may be be able to add simple keyword labels to individual trips “e.g. going to work”, or “grocery shopping”. If they do this daily, a large portion of data processing that is now the responsibility of a single analyst (and in this case, there were more than 50,000 trips in the original data) can be effectively and very cost-efficiently crowdsourced. Which means that an invaluable for the subsequent analysis contextual level of detail can be added to the data at a minimal small cost to the overall study.

— The effects of extreme weather, through the addition of high-resolution weather information in ArcGIS may also be interest. Again, resource constraint prevented this type of analysis during this study.