Capstone Data Project
BUS 4045
Harbourfront Volunteer Program

Assignment #3: Final Report

Alexei Marcilio 100988494
Client Sponsor: Karen Sinotte

Date: Thursday, April 7, 2016

Contents

List of Figures

Executive Summary

The Harbourfront Centre is an Arts and Cultural non-profit organization that would like to know more about their volunteers. Each year they have 450 volunteers from diverse backgrounds. It is time consuming training new volunteers so they would prefer volunteers who return year after year and also work more hours. They would like to know the profile of a typical volunteer. Also, they would like to know if certain demographic characteristics of volunteers correlated with greater tenure and more volunteer hours. An excel file of all volunteers which included many demographic variables as well as tenure and hours volunteered was examined to answer these questions. Most volunteers are female students. However older volunteers tend to work longer hours and return year after year. In addition non-students, and those with more than one language work significantly longer hours and have greater tenure. Those who live closer to Harbourfront work longer hours, but have no greater tenure. Four clusters of volunteers were identified. Ways in which to encourage more elderly volunteers were suggested, including making the program known at local retirement homes. Motivation strategies based on the four clusters identified were also discussed.

Problem Definition and Approach to Problem

Introduction

- Harbourfront Centre is an Arts & Cultural non-profit organization established in 1972
    - Over 4,000 events annually
    - 16 million visitors each year
    - Over 7,000 volunteer shifts
- Over 450 Volunteers Annually
- Diverse group of Volunteers – different ethnicities, ages, and professions

Challenges

- On-boarding of new volunteers is time consuming and costly
- Harbourfront Centre would like to better understand volunteer motivation and demographics in order to increase repeat volunteers and volunteer hours

Data Available

- A complete file of all 2015 Volunteers was made available
    - Includes data on 1269 volunteers

    - Includes demographic data, age, sex, school, language, employment status, and postal code
    - Includes tenure (we can assume that > 1 year is a returning volunteer
    - Includes volunteer hours
    - A great deal of clean up was done on the data – city and distance from Harbourfront Centre were calculated for each volunteer using Canada Post tools

- A file of Survey Monkey survey results was made available
    - Includes 110 respondents
    - 45 questions, including 6 demographic ones – age, sex, employment status, student or not, and IP Address
    - IP Address was used to determine city in order to perhaps correlate this to responses

Research Questions

- What does a typical volunteer "look like"?
    - Using visualization tools to graphically examine the profile of volunteers

- Do volunteers who return, and those who work more hours correlate with certain demographics?

- What do the survey results reveal about the motivation of volunteers and their satisfaction with the volunteer program:

    - The survey data will be visualized in Tableau in order to draw conclusions

Approach

In order to answer these research questions I did the following:

1. Cleaned up the data when necessary
2. Validated all postal codes and added "Distance to Harbourfront" as a measure using Canada Post's lookup service
3. Visualized the data in Tableau which would allow for me to answer questions concerning the typical volunteer profile
4. T-Tests to determine if Hours volunteered and Tenure are different for men and women, older vs younger, English speakers only vs those with another language, and those who live close vs those who live farther away
5. Performed Segmentation Analysis and Decision Tree Analysis to segment the data into different groups using SAS Data Miner

Results

Overview of Volunteers

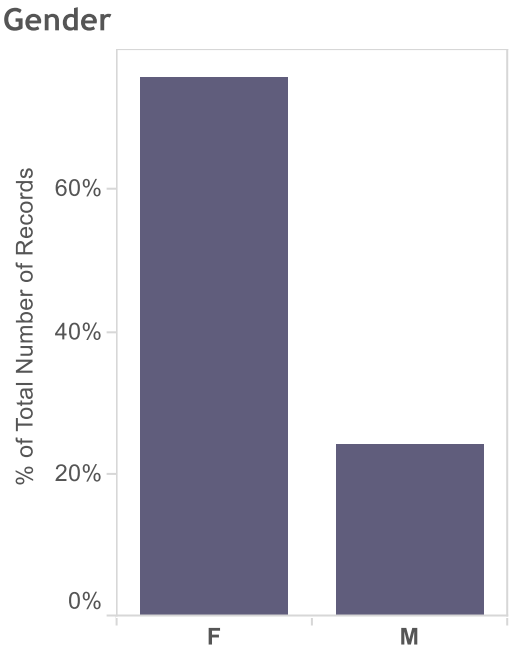


Figure 1. Gender makeup of volunteers

Most of the volunteers are female at about 75%.

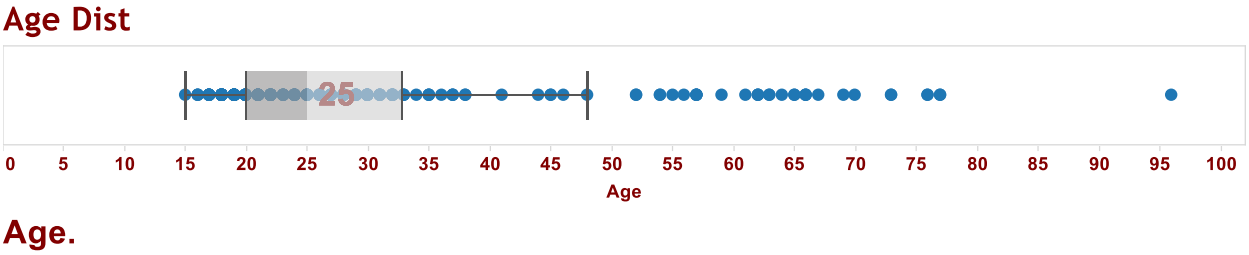


Figure 2. Age Distribution of Volunteers.

The average age of a volunteer is 25 however this is heavily skewed to the right. This makes sense as there are a great deal of student volunteers. The youngest volunteer is 15 and the oldest is 96.

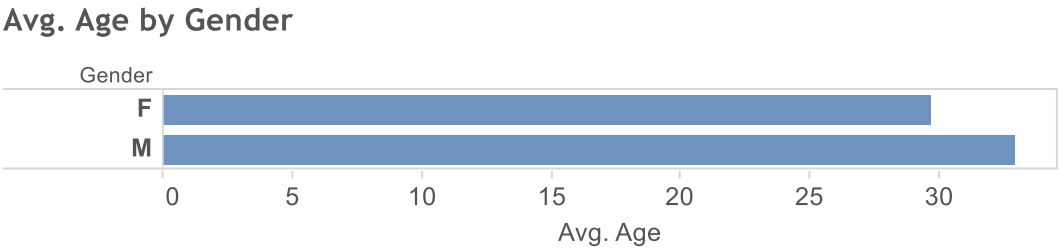
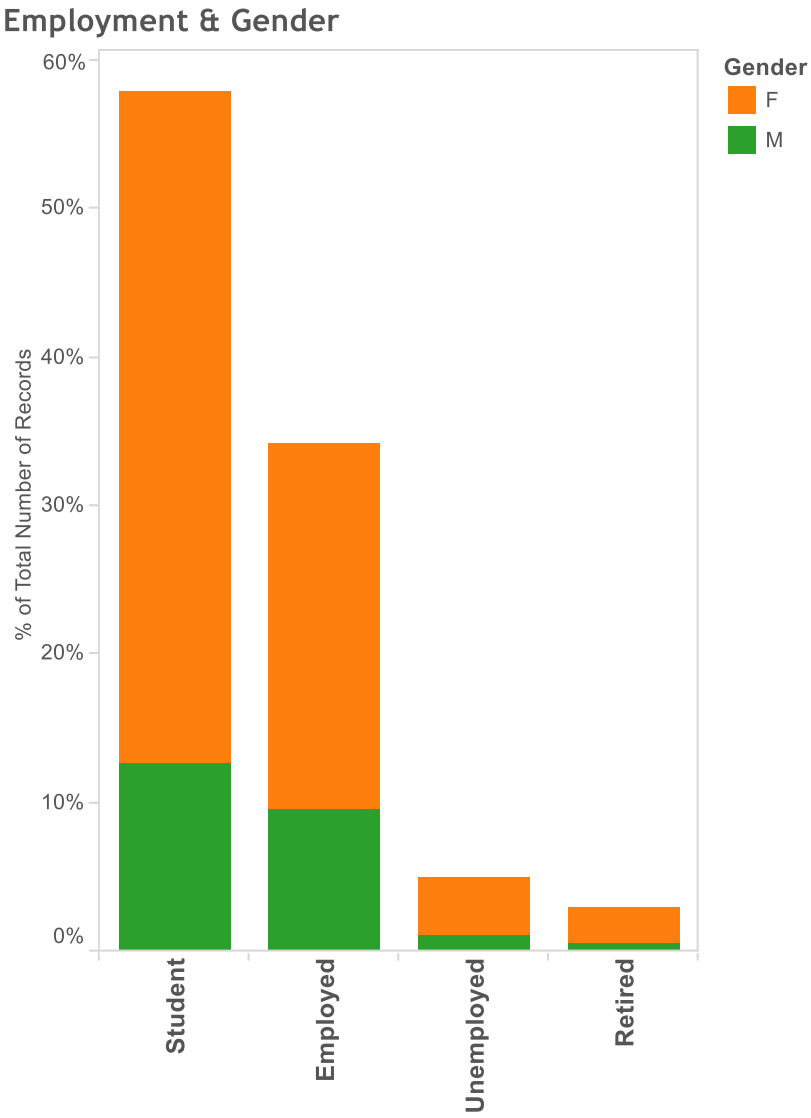


Figure 3. Age by Gender of Volunteers.

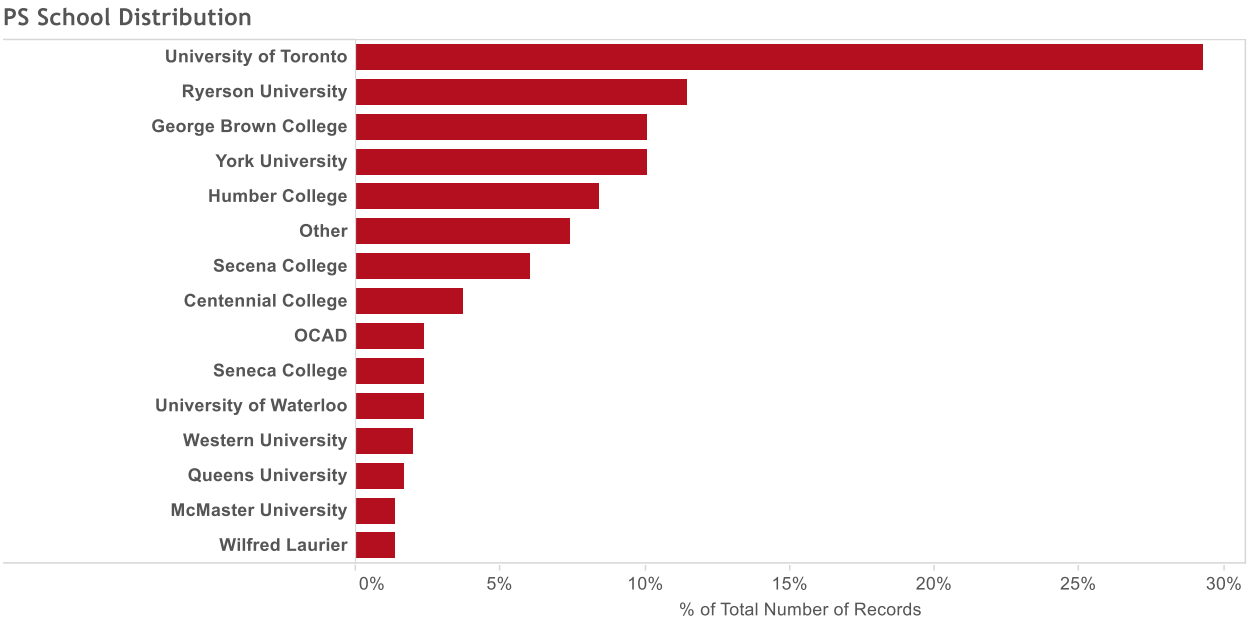
There's not much different in the ages based on gender.



% of Total Number of Records for each Employment Status (group). Color shows details about Gender. The view is filtered on Employment Status (group) and Gender. The Employment Status (group) filter keeps Employed, Retired, Student and Un-employed. The Gender filter keeps F and M. Percents are based on the whole table .

Figure 4. Employment and Gender of Volunteers.

However it does seem that males make up a much smaller percentage of Unemployed and Retired volunteers. Students make up the largest proportion of volunteers at about 58%. Retired people make up the smaller percentage.



% of Total Number of Records for each School (group). The data is filtered on School, which excludes Null and High School. Percents are based on the whole table .

Figure 5. Distribution of Schools of Volunteers.

If we look at the distribution of schools The University of Toronto has by far the largest number of volunteers. More than twice as much as the next highest which is Ryerson, followed by George Brown College and York University. Could this be due to the fact that UofT is the largest University in Toronto? Or perhaps there are other factors. Maybe the Harbourfront Centre promoted the program at UofT?

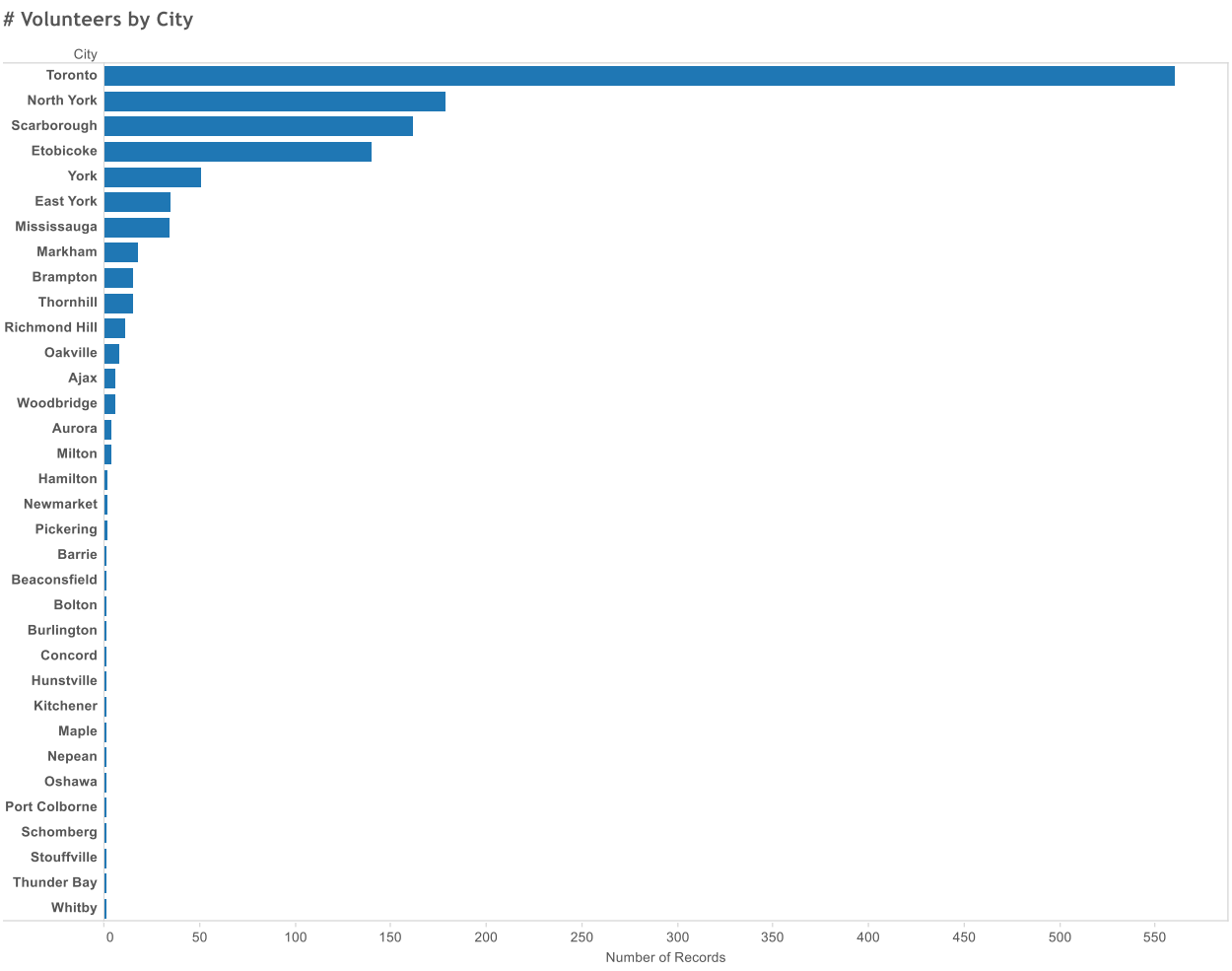


Figure 6. Distribution of Cities of Volunteers.

By far the largest number of volunteers come from Toronto. This makes sense as Harbourfront is in Toronto by the lake which makes it easiest to access from downtown Toronto.

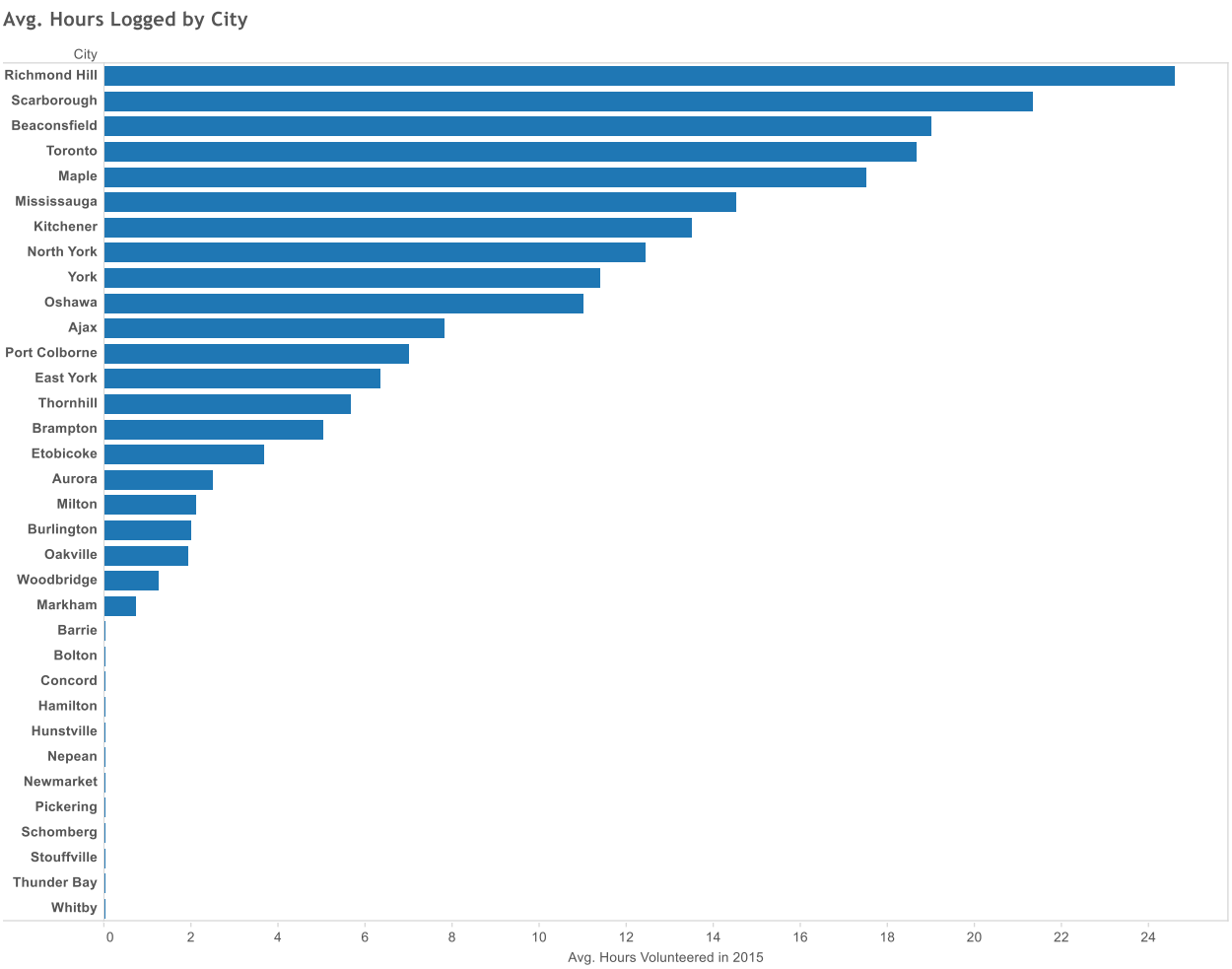
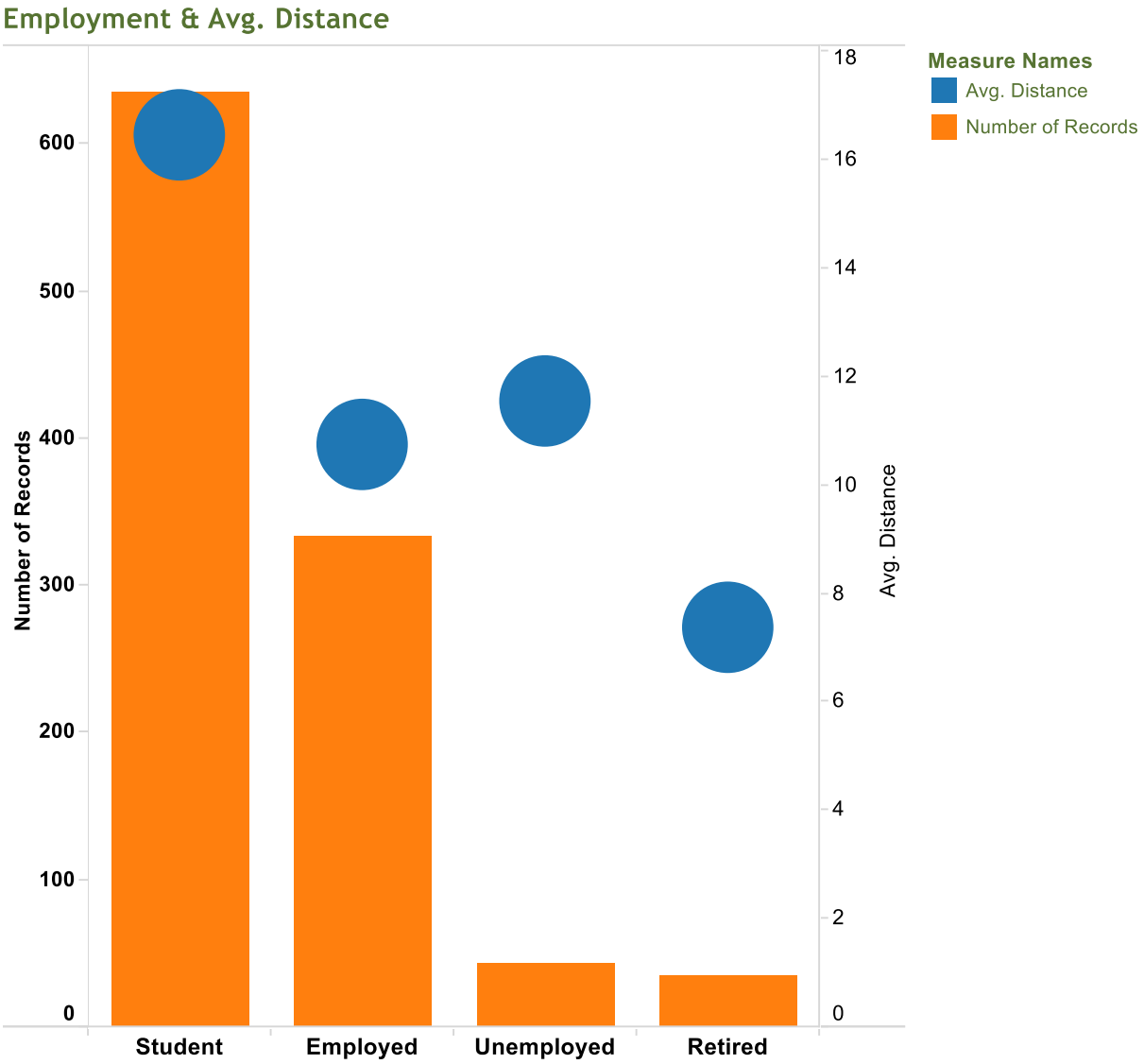


Figure 7. Volunteer Hours Logged by City.

However if we look at hours average hours logged by the city in which the volunteer lives Toronto is not the first, there are some suburban volunteers doing a great deal of volunteer work.



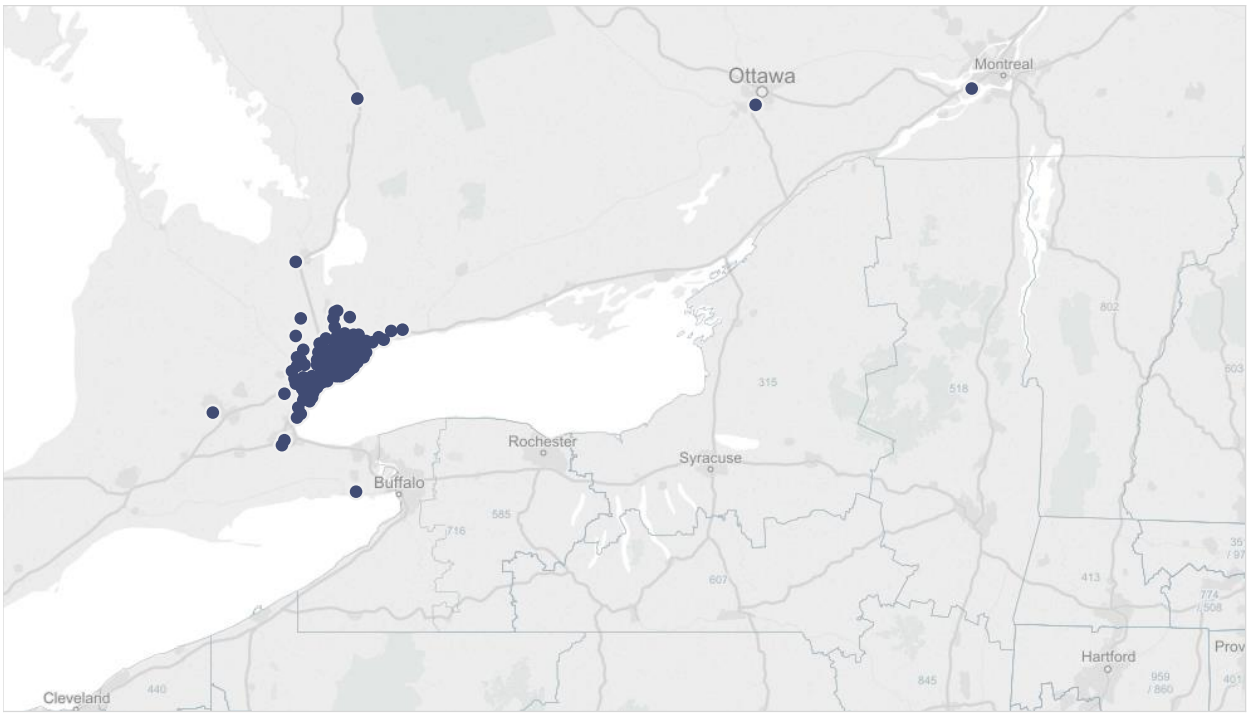
Number of Records and Avg. Distance for each Employment Status (group) 1. Color shows details about Number of Records and Avg. Distance. The view is filtered on Employment Status (group) 1, which keeps Employed, Retired, Student and Unemployed.

Figure 8. Employment and Distance of Volunteers.

We were able to measure the distance from Harbourfront to each volunteer's home by using a Canada post program that measures distance between two postal codes. The distance returned is not "by the crow flies" but instead a more meaning driving distance. This is more useful because places that are further away from Harbourfront may not necessarily be more difficult to get to and a straight line distance measurement is misleading.

In the graph above we can see that students make up the largest proportion of volunteers, and they are also the furthest away with an average distance of about 30% more than unemployed and employed. Retired volunteers are by far live the closest to Harbourfront.

Geographic Distn.

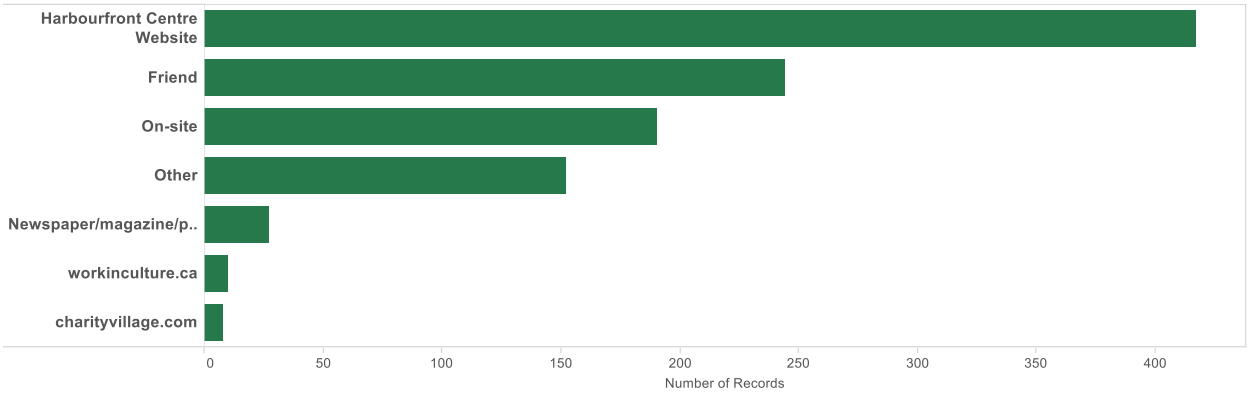


Map based on Longitude (generated) and Latitude (generated). Details are shown for FSA. The view is filtered on FSA, which excludes P7B.

Figure 9. High Level Map of Volunteer's homes.

Here's a high level plot of each volunteer on a map of Ontario. Not surprising most volunteers live around the GTA but there are a few outliers.

How did Your Hear?



Sum of Number of Records for each CF - Where did you hear about us?. The view is filtered on CF - Where did you hear about us?, which excludes Null.

Figure 10. How Did You Hear About Harbourfront?

When asked how they first heard of Harbourfront Centre most volunteers mentioned the Website, followed by friends. The fact that many volunteers heard of the program "on-site" might contribute to a bias favoring more local people, however it is a popular site for suburban people to visit.

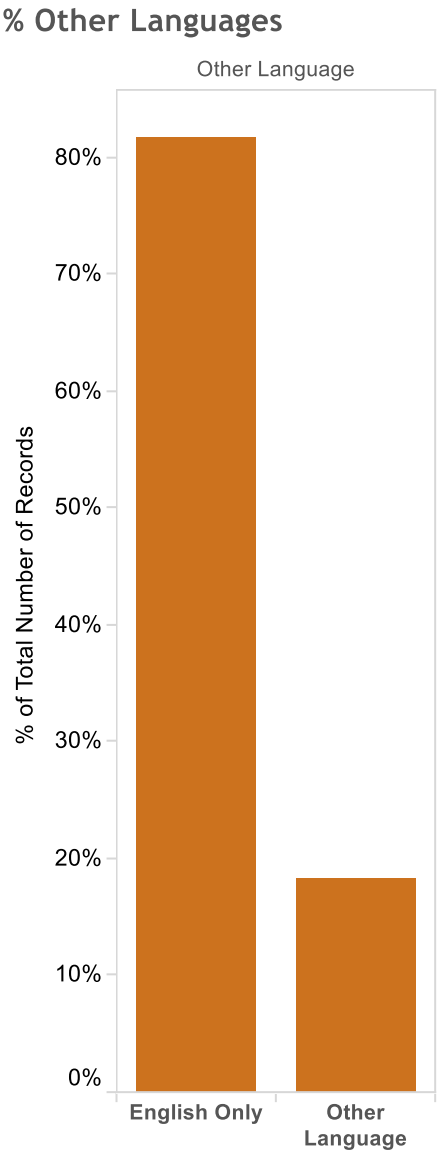


Figure 11. Other Languages Frequency.

Let's look at languages. The majority of volunteers are English only speaking as less than 20% indicate another language. It could be however that some prefer not to disclose this.

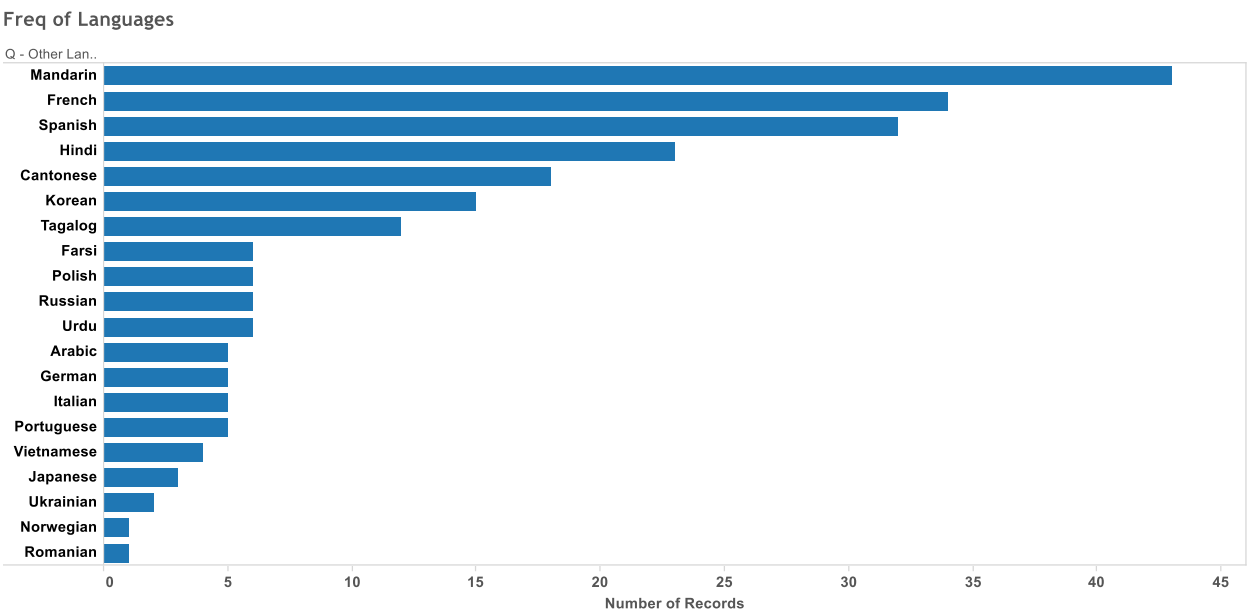
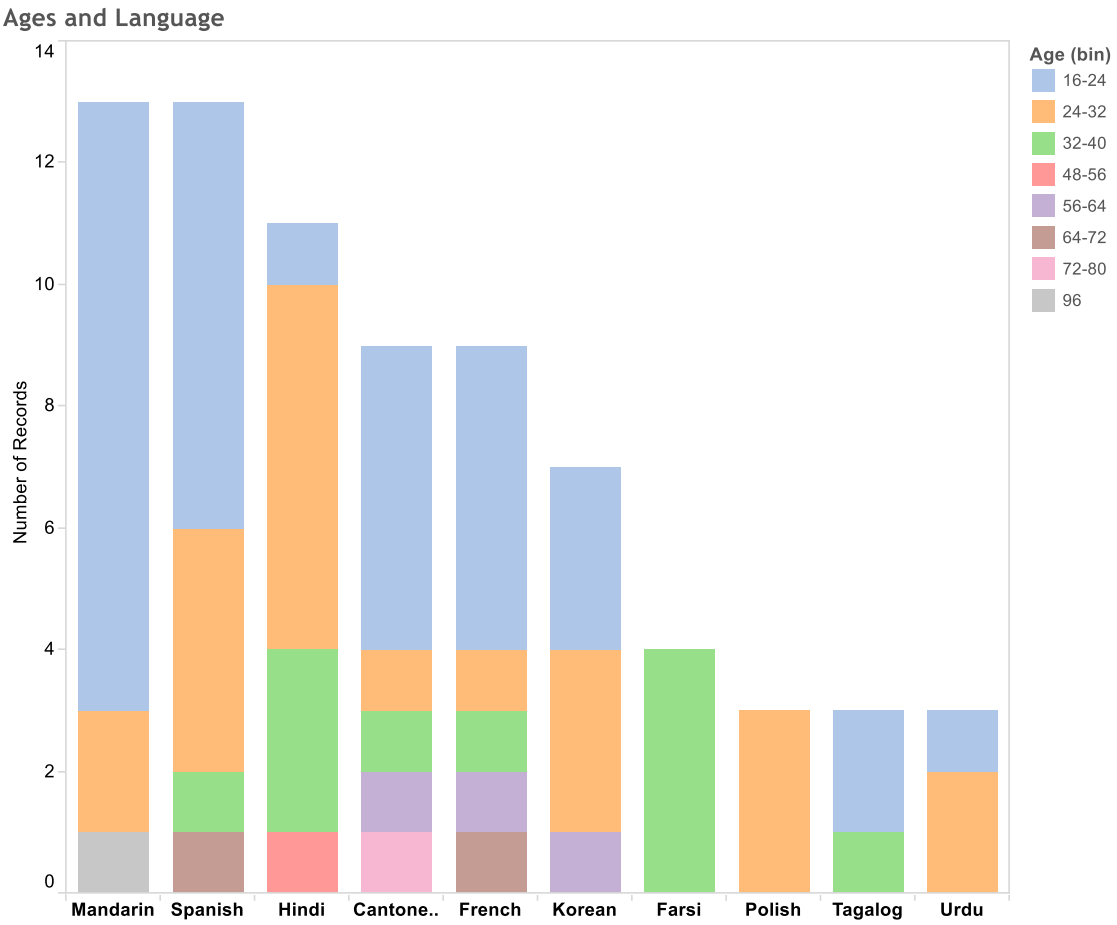


Figure 12. Frequency of Languages.

If we look at all the volunteers that indicated that they speak another language we can see that Mandarin is the most spoken, followed by French, Spanish and Hindi.

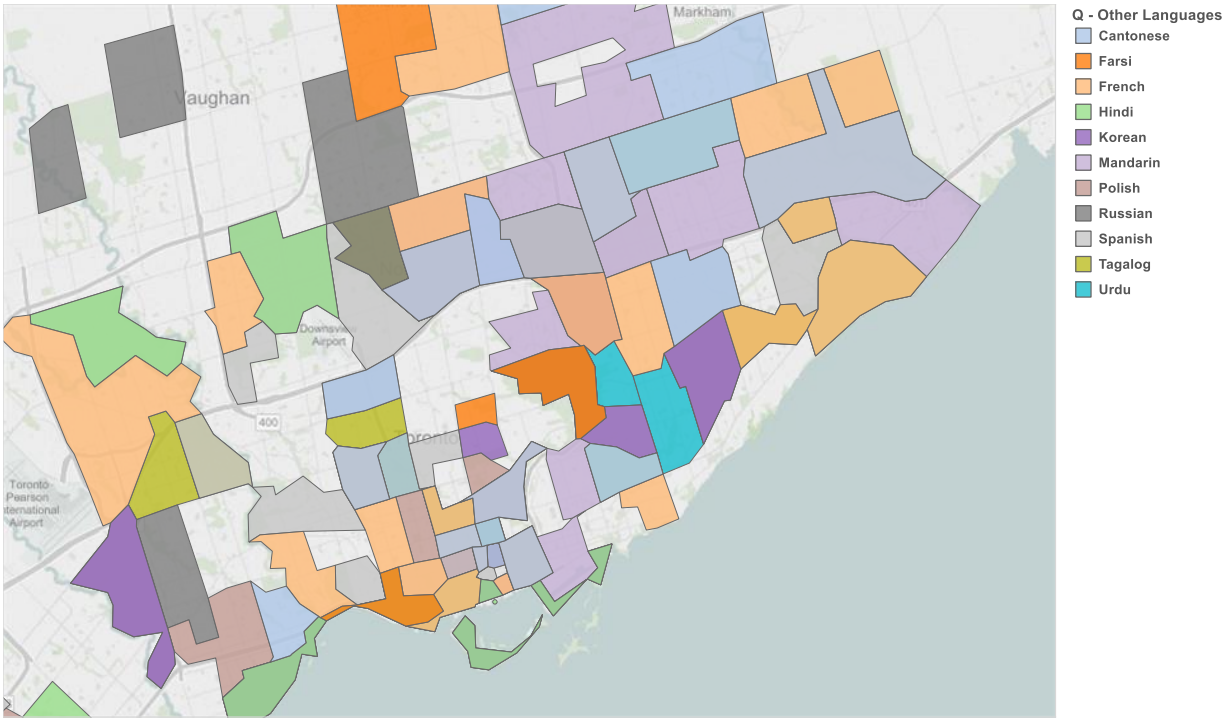


Sum of Number of Records for each Q - Other Languages. Color shows details about Age (bin). The view is filtered on Q - Other Languages and Age (bin). The Q - Other Languages filter keeps 11 of 21 members. The Age (bin) filter excludes Null.

Figure 13. Ages and Languages.

It's also interesting that Mandarin speakers are the youngest average age group with a large proportion, almost 80% being 16-24 in age.

Distribution of Languages



Map based on Longitude (generated) and Latitude (generated). Color shows details about Q - Other Languages. Details are shown for FSA. The view is filtered on Q - Other Languages, which keeps 11 of 21 members.

Figure 14. Distribution of Languages in the GTA.

If we look at the distribution of languages in the GTA we can see there is a pattern. French and Spanish seem more widespread, Hindi speakers are more in the east side of the city and Farsi speakers are more central. Cantonese speakers are widespread with a concentration downtown.

What Correlates with Tenure and Hours Worked?

Now let's examine some correlations to see if inferences can be made about tenure and total number of hours worked.

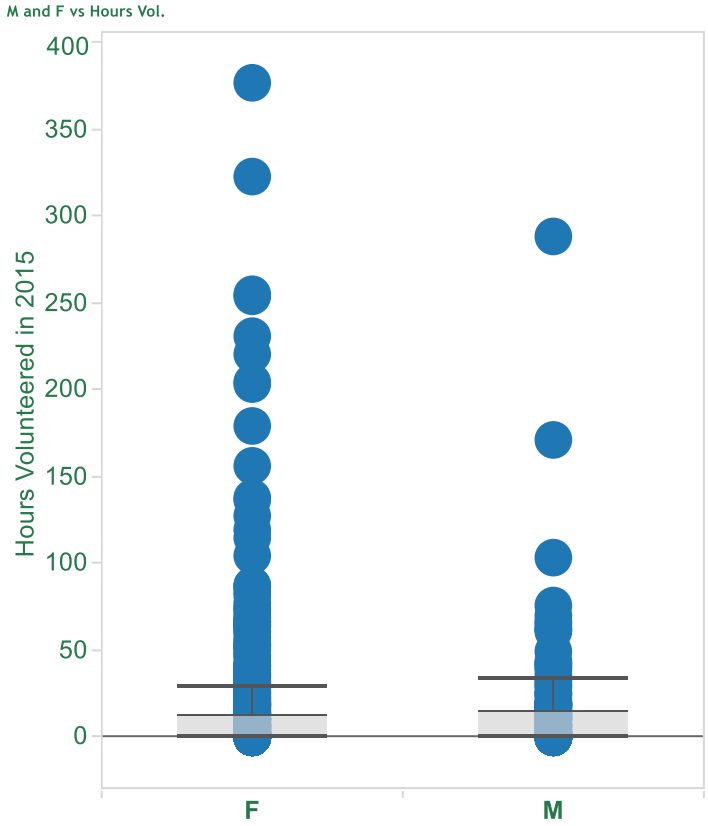
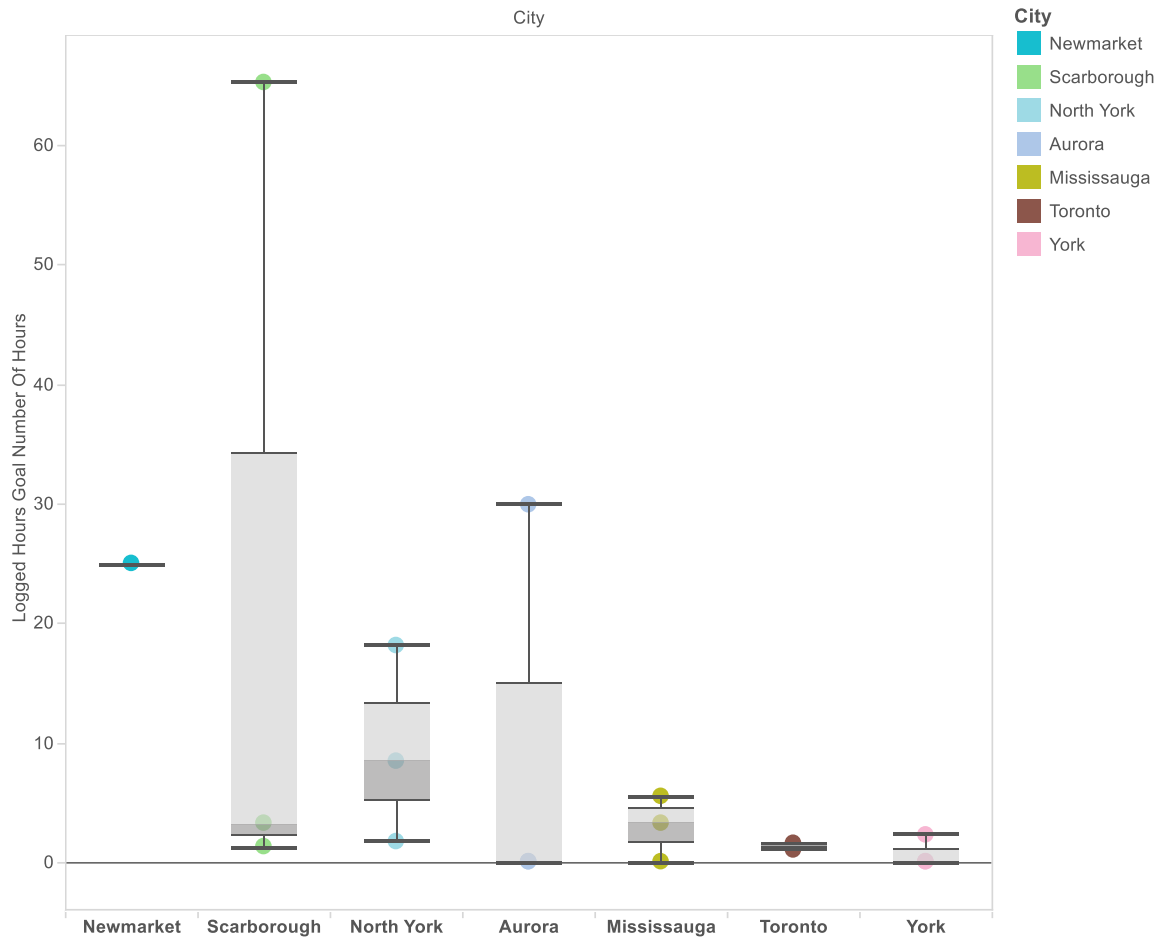


Figure 15. Hours worked Men vs Women.

It appears that there are no statistical difference for hours worked between men and women.

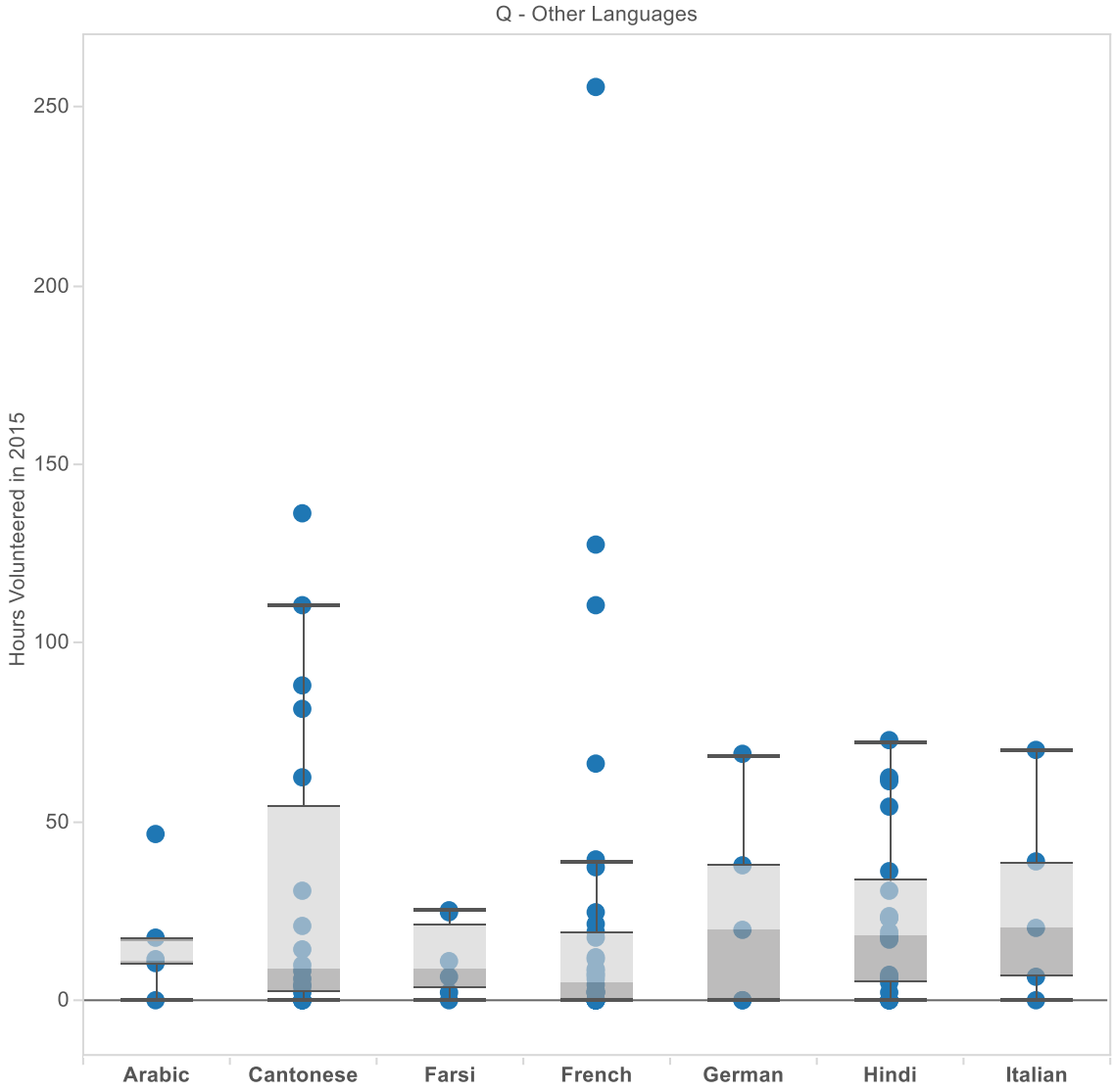


Average of Logged Hours Goal Number Of Hours for each City. Color shows details about City. Details are shown for Gender. The view is filtered on City and average of Logged Hours Goal Number Of Hours. The City filter keeps 7 of 34 members. The average of Logged Hours Goal Number Of Hours filter keeps all values.

Figure 16. Hours logged by City.

If we examine hours worked by city there's a large range in Scarborough and Aurora, but not so in other cities. Toronto seems suspiciously narrow so it might be wise to re-examine this again.

Language vs. Hours



Hours Volunteered in 2015 for each Q - Other Languages. The view is filtered on Q - Other Languages, which keeps 7 of 21 members.

Figure 17. Hours logged by Language.

Likewise there seems to be no relationship between languages spoken and hours worked.

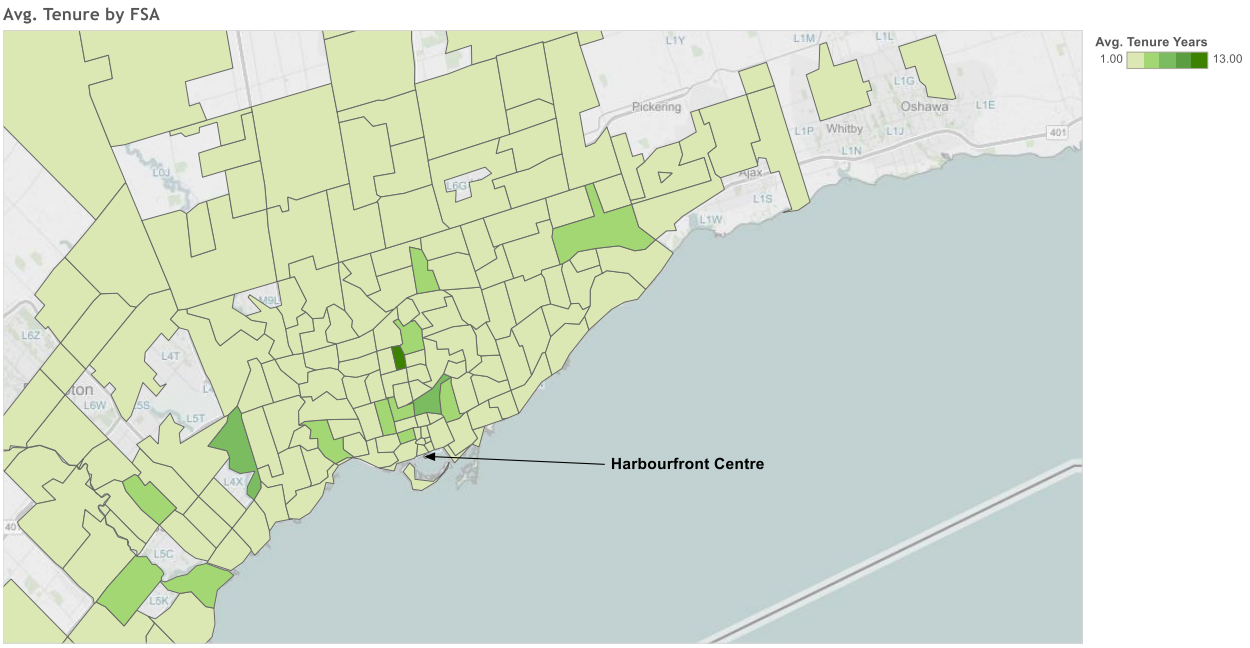


Figure 18. Tenure Map of GTA.

Let's examine distance from Harbourfront to see if it is a factor. Here's a plot of average tenure for each FSA in the GTA. Perhaps we can see a pattern. The greener FSAs are located closer to Harbourfront. Even though there are some further away, the fact they are close to the lake might make it easier to get into Harbourfront along the lakeshore highway for example.

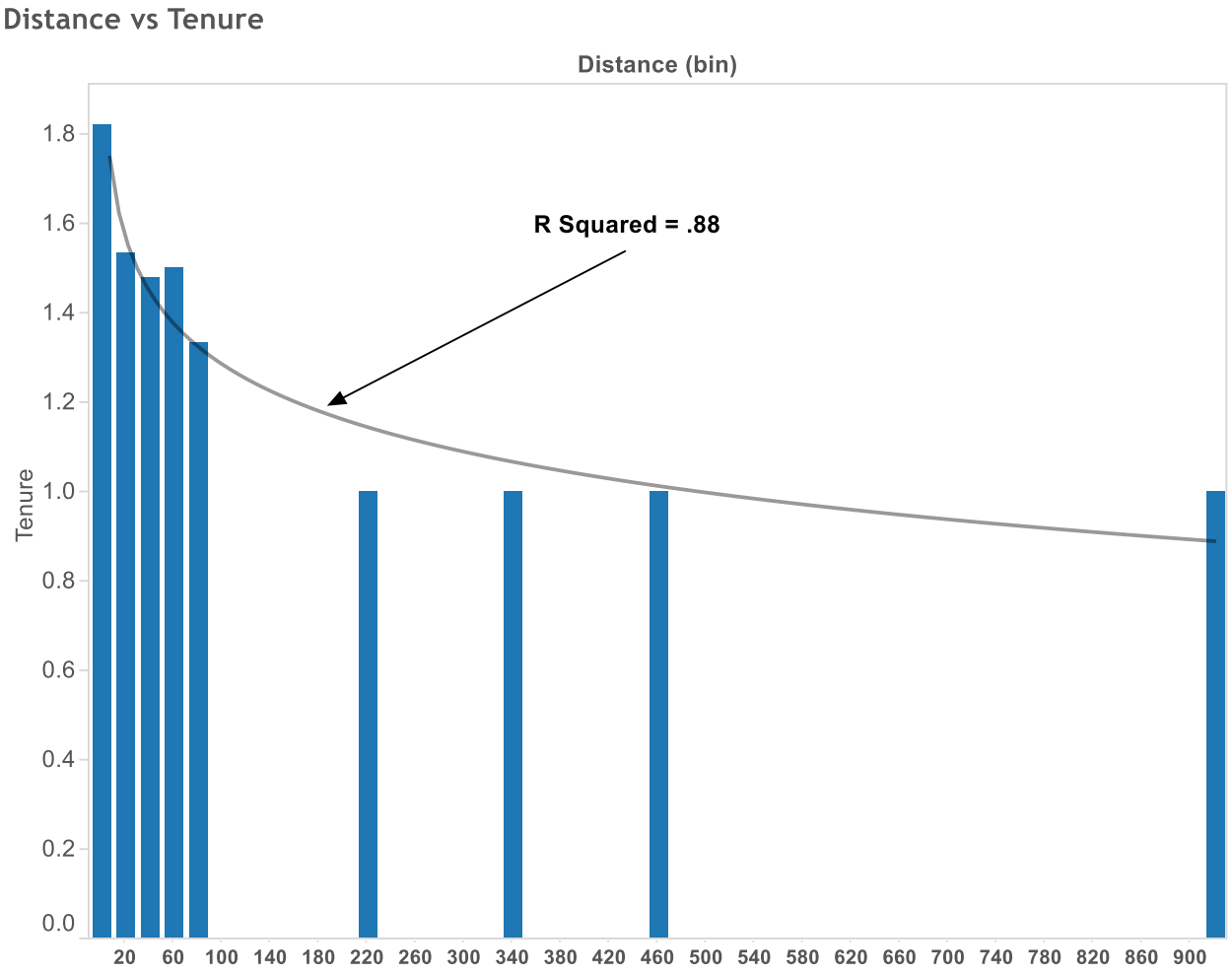
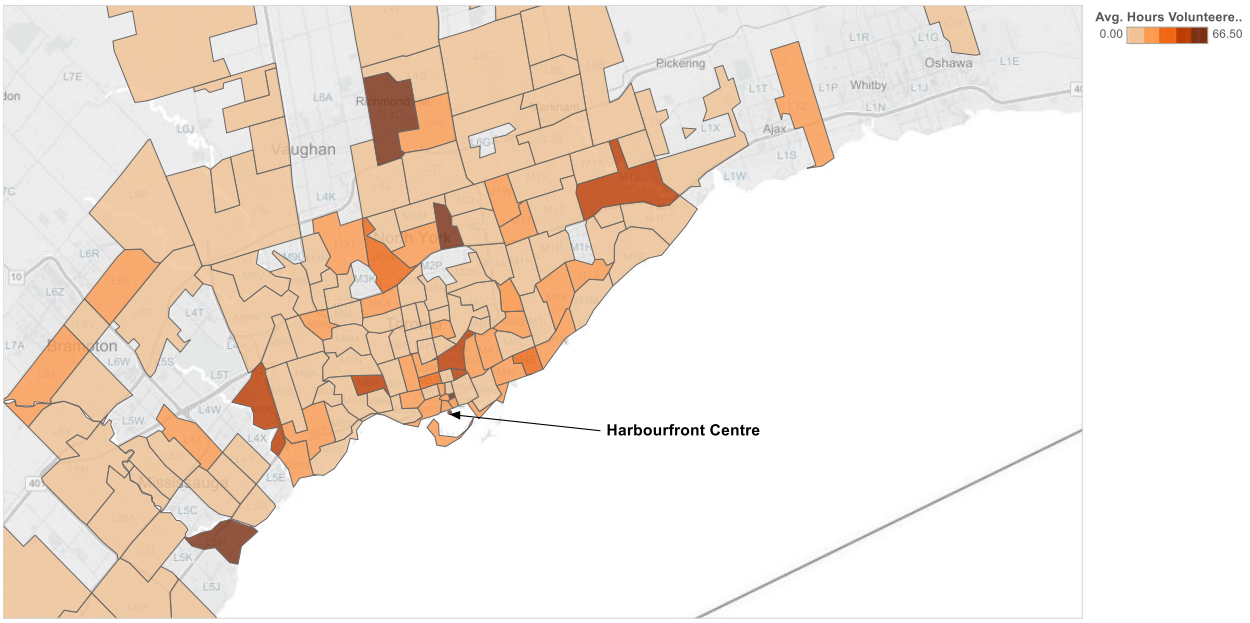


Figure 19. Distance vs Log Tenure.

Here I have grouped volunteers by distance in bins of 20 KM. This is plotted against Tenure. We can see that a logarithmic curve fits the data quite well with an R Squared of .88 which is high. Distance from Harbourfront seems to correlate with tenure.

Avg. Hours Volunteered

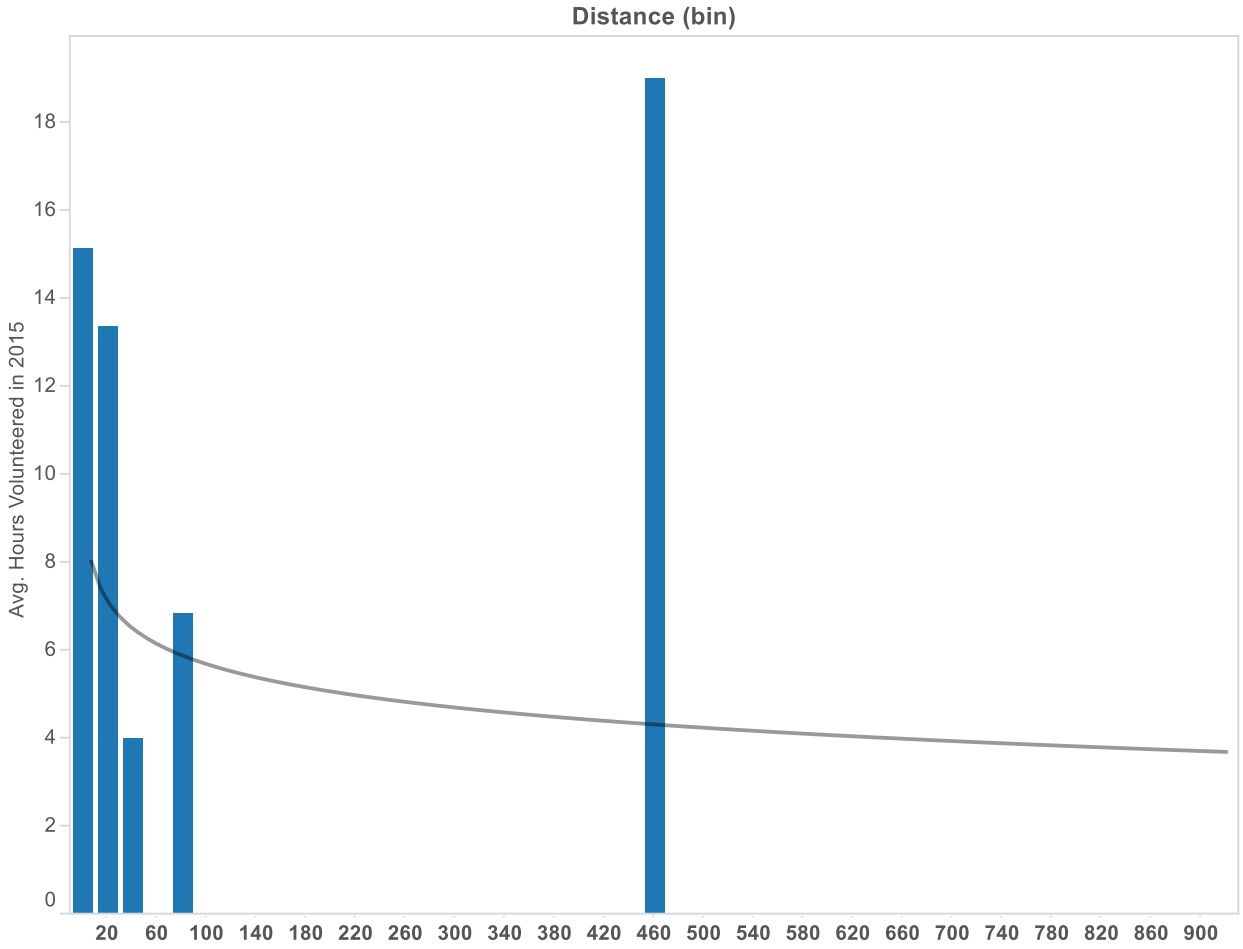


Map based on Longitude (generated) and Latitude (generated). Color shows average of Hours Volunteered in 2015. Details are shown for FSA. The data is filtered on Gender, which keeps F and M. The view is filtered on FSA, which excludes P7B.

Figure 20. Hours worked by FSA.

Let's look at a plot of average hours worked by FSA. I am not sure there is a pattern, but perhaps there is a weaker one.

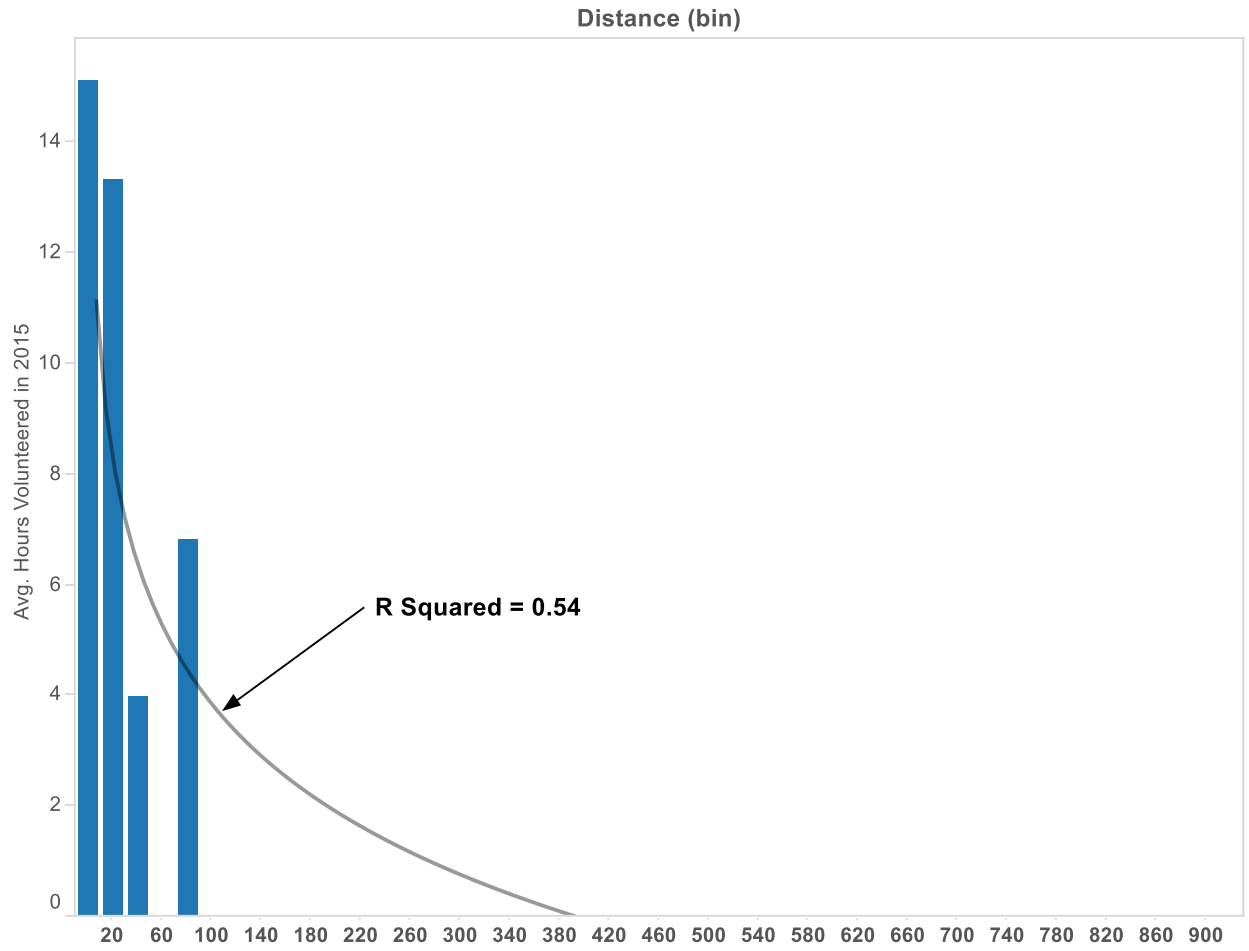
Distance by Avg. Hours



Average of Hours Volunteered in 2015 for each Distance (bin). The view is filtered on Distance (bin), which excludes Null.

Figure 21. Distance by Average Hours.

It looks like there is a definite outlier if we plot bins of distance against average hours volunteered.

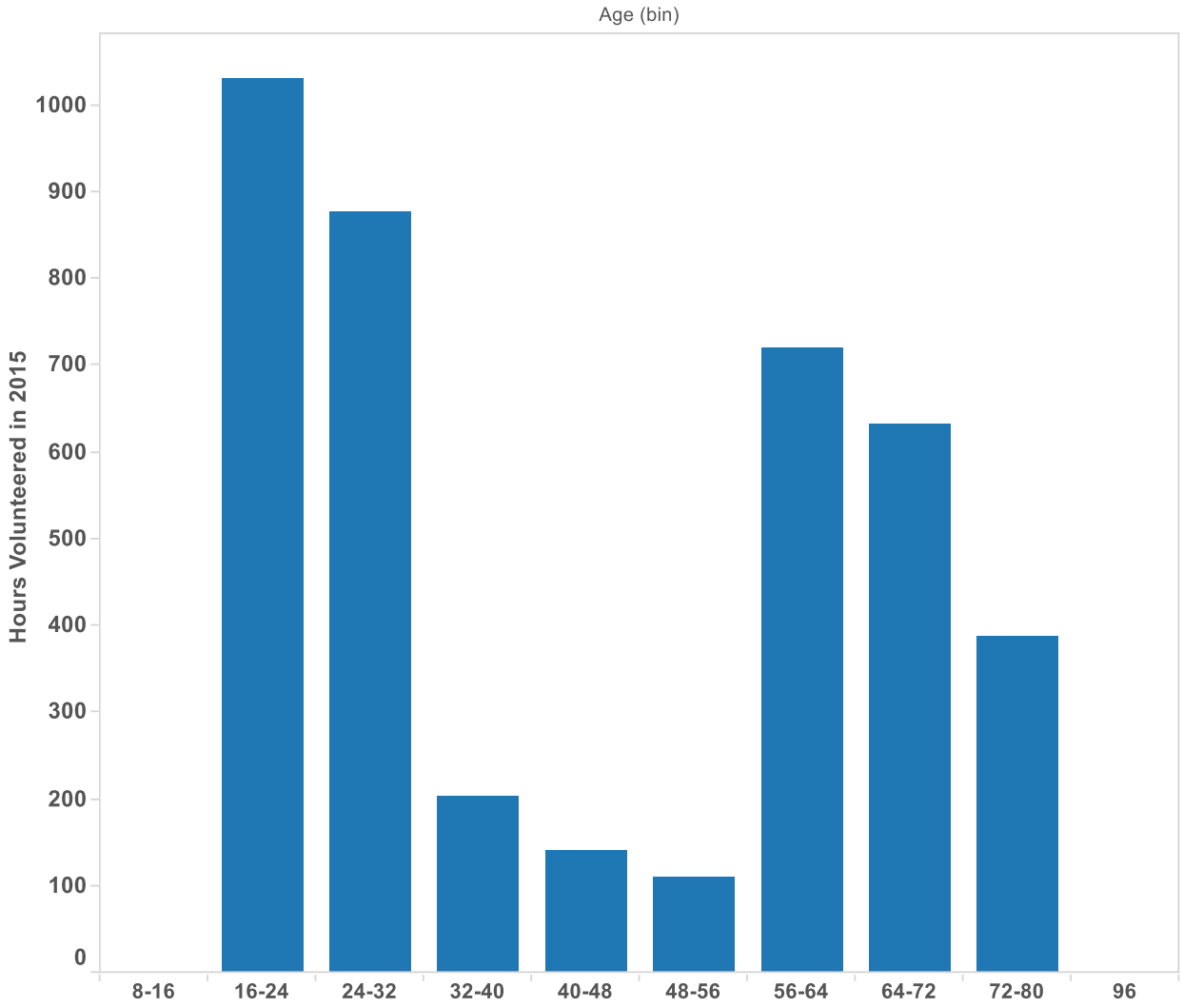


Average of Hours Volunteered in 2015 for each Distance (bin). The view is filtered on Distance (bin), which excludes Null and 460.

Figure 22. Distance by Average Hours minus outlier.

If we remove this outlier we can see some of the variation is explained and there does seem to be a relationship with distance and average hours worked.

Age Group and Total Hours



Sum of Hours Volunteered in 2015 for each Age (bin). The view is filtered on Age (bin), which excludes Null.

Figure 23. Age Group and total hours.

Now let us examine the age of the volunteers and how that relates to hours worked. If we look at the total number of hours worked people from 16-32 followed by those from 56-80 make up the largest groups.

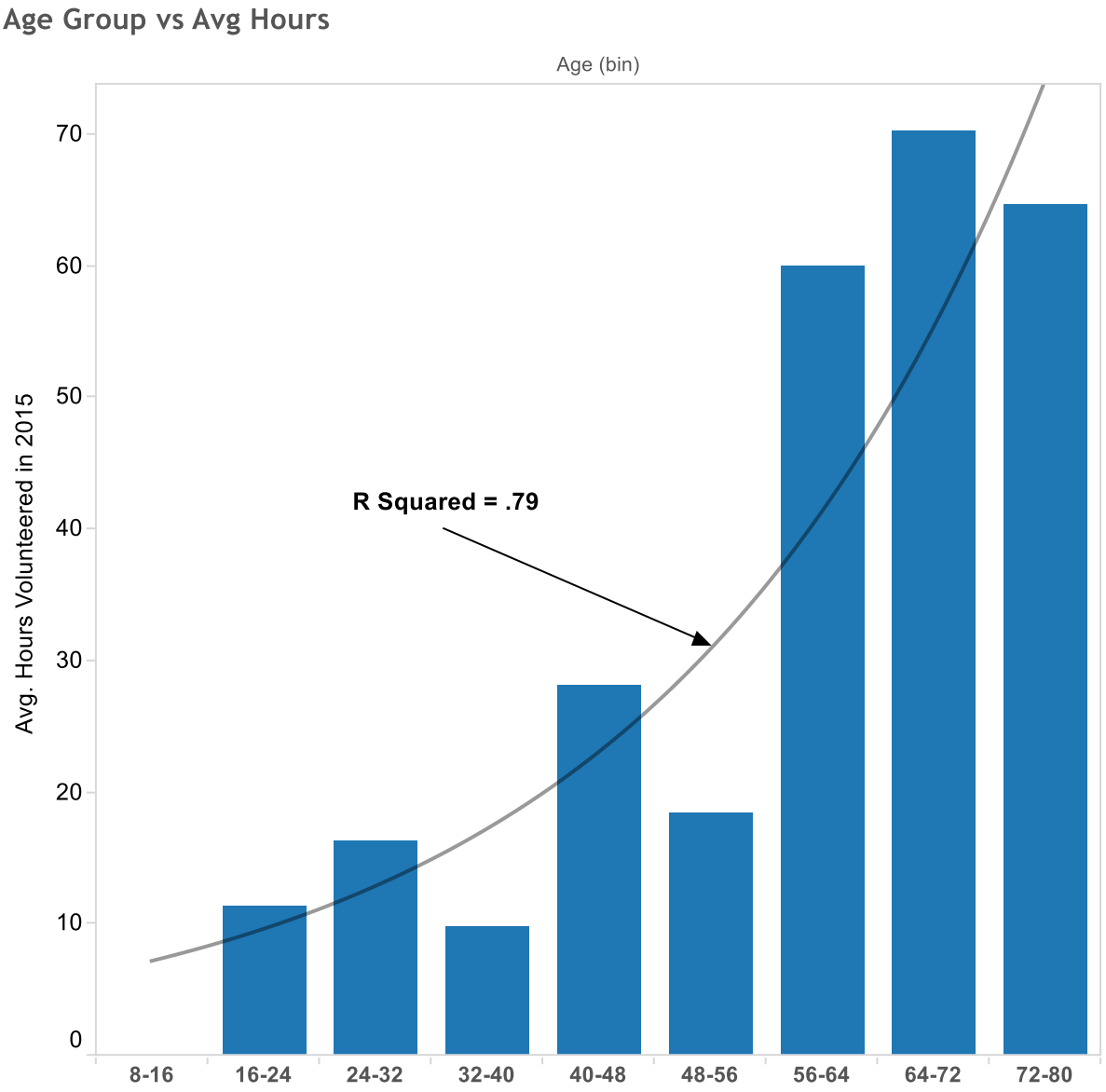


Figure 24. Age Group vs Average Hours.

If instead we look at average hours volunteered by age group a pattern emerges. The older a volunteer is the more average hours he or she works. The R squared of a log curve is .79 which is quite high.

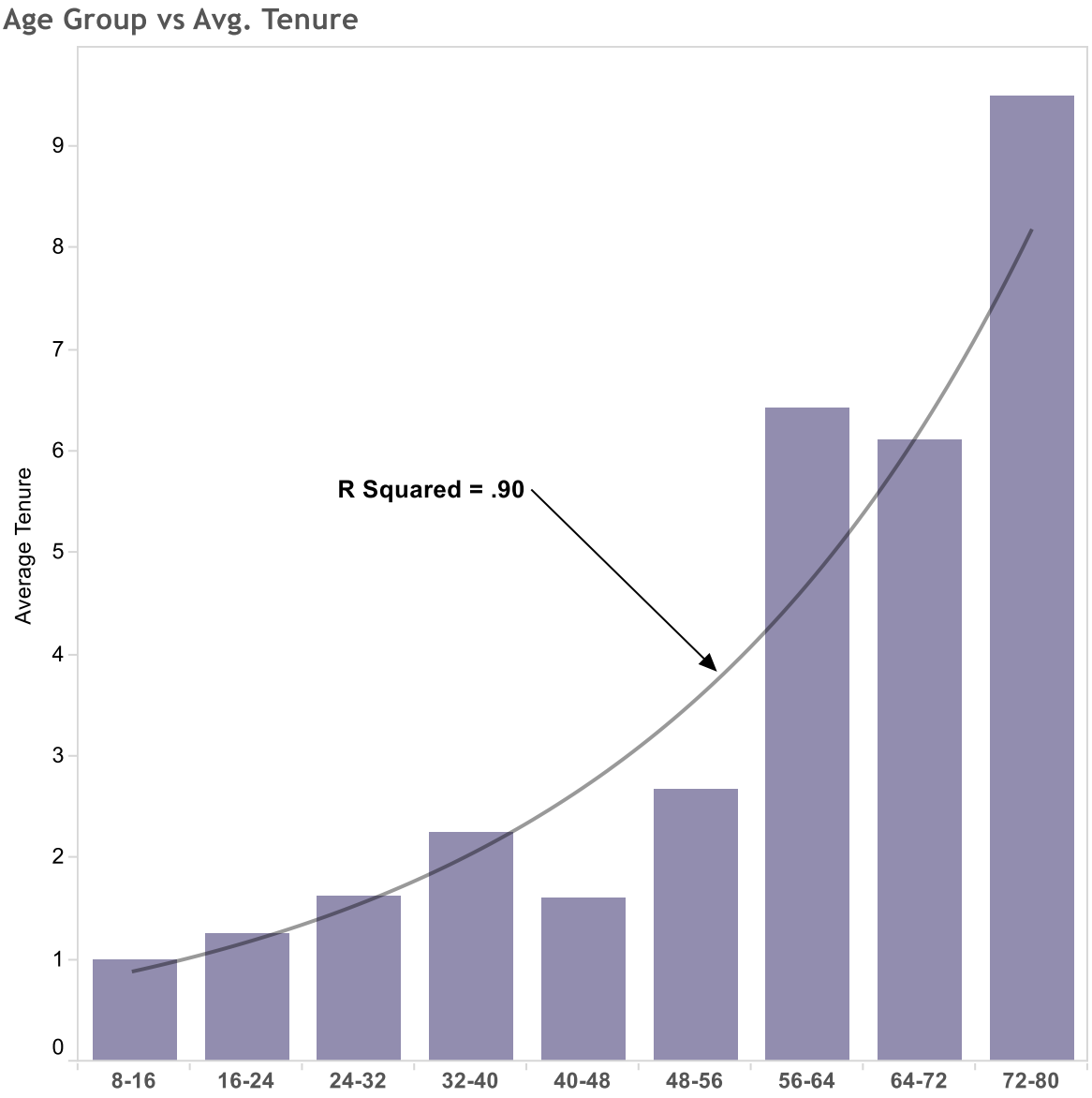


Figure 25. Age Group vs Average Tenure.

An even stronger relationship emerges when we look at average tenure vs age group. The R Squared for a logarithmic curve is .9 which is very high.

T-Tests – Differences in the Means

T-Tests assuming unequal variances were used to test the difference between mean tenure and mean hours worked for older vs younger (< 30 vs > 30), closer living vs further away (<2 km vs > 2 km), male vs female, English only speakers vs those with other languages and students vs non-students.

The results are as follows:

Hours Volunteered

Mean		P-Value	Significant?
Male	Female		
14.90	14.27	0.811	
English	Other		
13.63	19.01	0.047	✓
Non-Student	Student		
20.34	8.90	0.000	✓
Age >=30	Age < 30		
32.50	12.87	0.004	✓
< 2KM	> 2 KM		
18.41	13.62	0.049	✓

Figure 26. T-Test results of Hours Volunteered.

There are significant differences between mean hours volunteered among English only Speakers and Multi Language speakers, students and non-students, older and younger volunteers and those who live close vs those who live further.

Tenure

Mean		P-Value	Significant?
Male	Female		
1.87	1.67	0.242	
English	Other		
1.64	2.39	0.004	✓
Non-Student	Student		
2.28	1.27	0.000	✓
Age >=30	Age < 30		
1.83	1.33	0.000	✓
< 2KM	> 2 KM		
1.95	1.73	0.235	

Figure 27. T-test results of Tenure.

There are significant differences between average tenure among English only Speakers and Multi Language speakers, students and non-students and older and younger.

Cluster Analysis

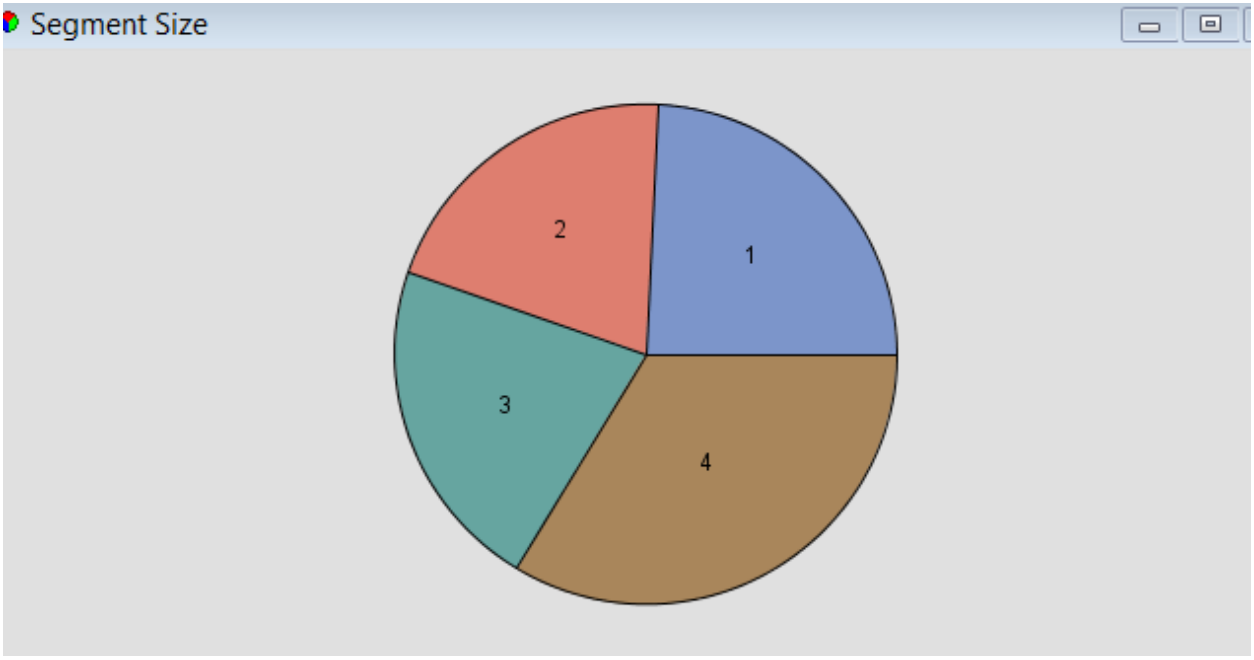


Figure 28. SAS Data Miner Segments.

I get four pretty even clusters. Again, I did rejected quite a few variables. I left Age, Distance, Employee Status, Gender and Other Languages. I thought these were all variables we would easily know when we first met the volunteers.

I made Gender and Student binary variables (1 for male, 0 for female) and 1 for Student, 0 for Not Student.

Mean Statistics

Segment Id	Frequency of Cluster	Age	Student	Distance	Gender
1	380	21.07595	0.95	17.74256	-6.1E-16
2	178	24.25926	0.382022	19.04781	1
3	277	57.9	0.072202	3.398303	1
4	355	38.22449	0.405634	5.523051	8.05E-16

Figure 29. SAS Data Miner Mean Statistics.

Groups:

1. Young Students Female
2. Young Males, mostly not students
3. Old Males Not Students
4. Middle Age Females – about 41% students

Figure 30. An imaginary rendering of each segment.

Decision Tree Analysis

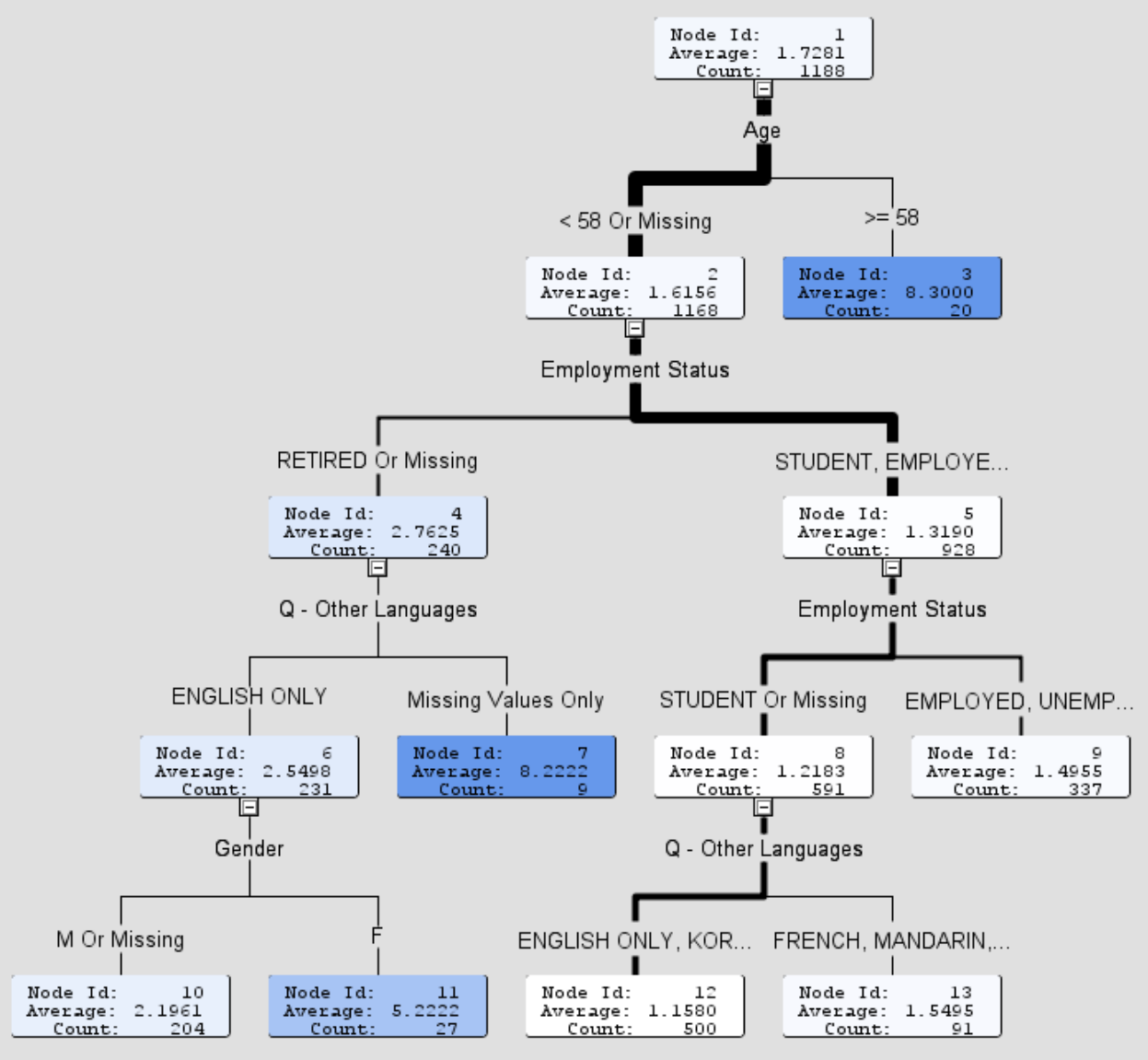


Figure 31. SAS Data Miner Decision Tree Analysis.

I used tenure as the target variable and excluded hours volunteered in 2015 because I assumed using the number of hours someone works is not a good predictor if we are just hiring them because we would not know this.

It looks like AGE is the largest decision criterion. Greater than 58 years old and younger than 58 years old. Seems to be a huge difference in the average tenure here – 8.3 years vs 1.6 years. The next level is based on employment status, with student, employed and unemployed vs those who are retired.

Survey Results: Satisfaction and Motivation

Let's look at the survey results.

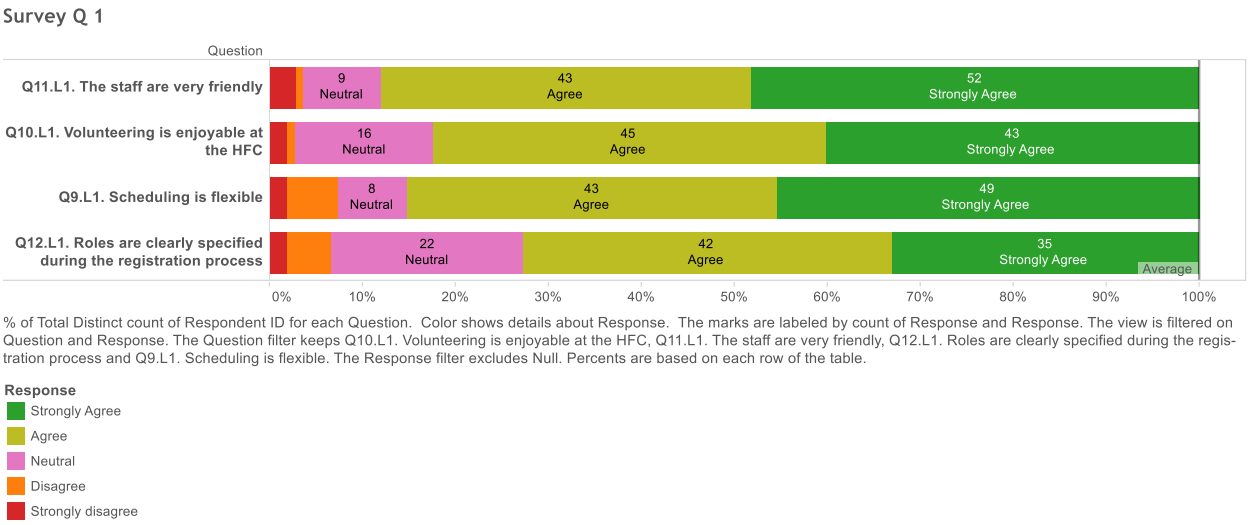


Figure 32. Survey Results Motivation Questions.

Overall satisfaction seems high with the vast majority of volunteers stating that they are happy with the staff, that they find volunteering enjoyable, that scheduling is flexible and the roles are clearly specified. The most negative response, which is not that negative at all, comes when they are asked if scheduling is flexible.

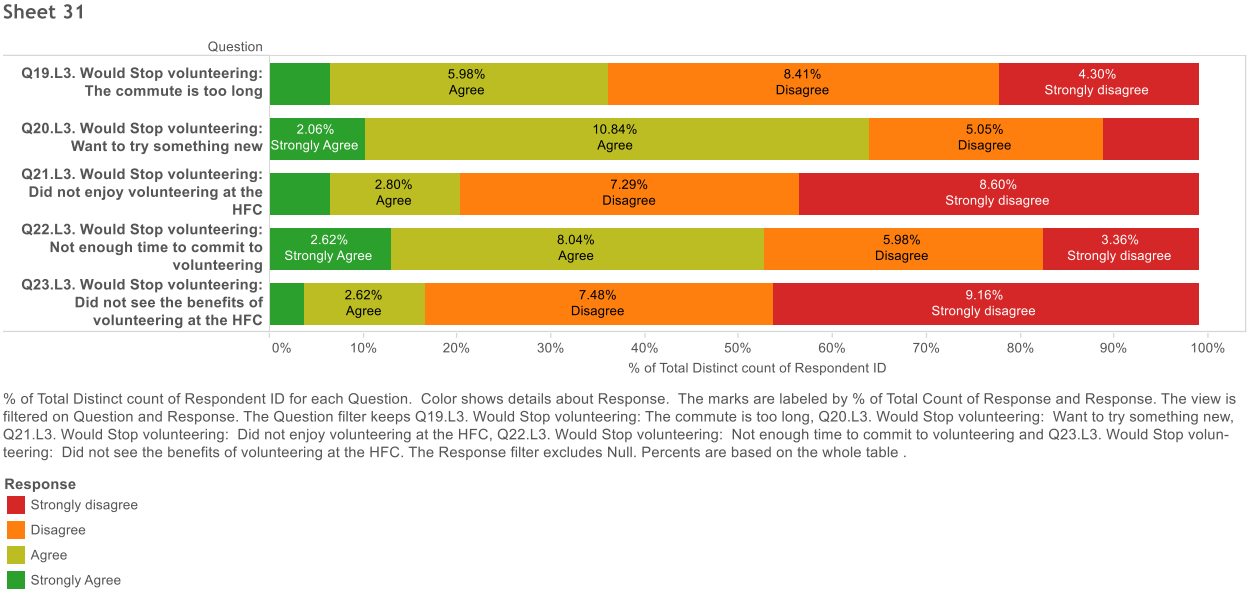


Figure 33. Survey Results - What would stop you?

Here we can see the responses when asked reasons why they would stop volunteering. The major reason seems to be just wanting to try something new.

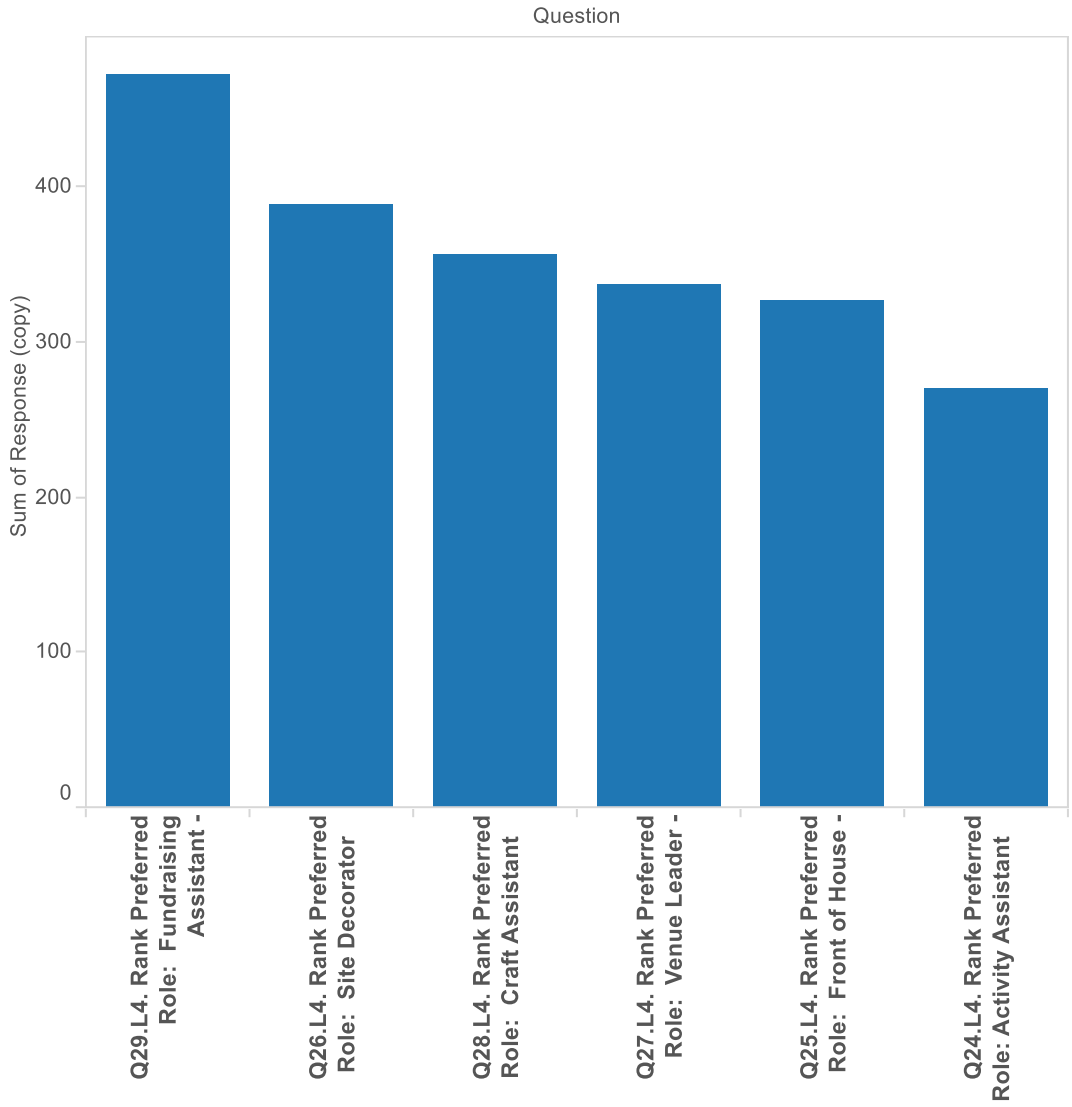


Figure 34. Survey Results Preferred Role.

When asked their preferred role it is Fundraising Assistant followed by Site Decorator. There's nothing that really sticks out here.

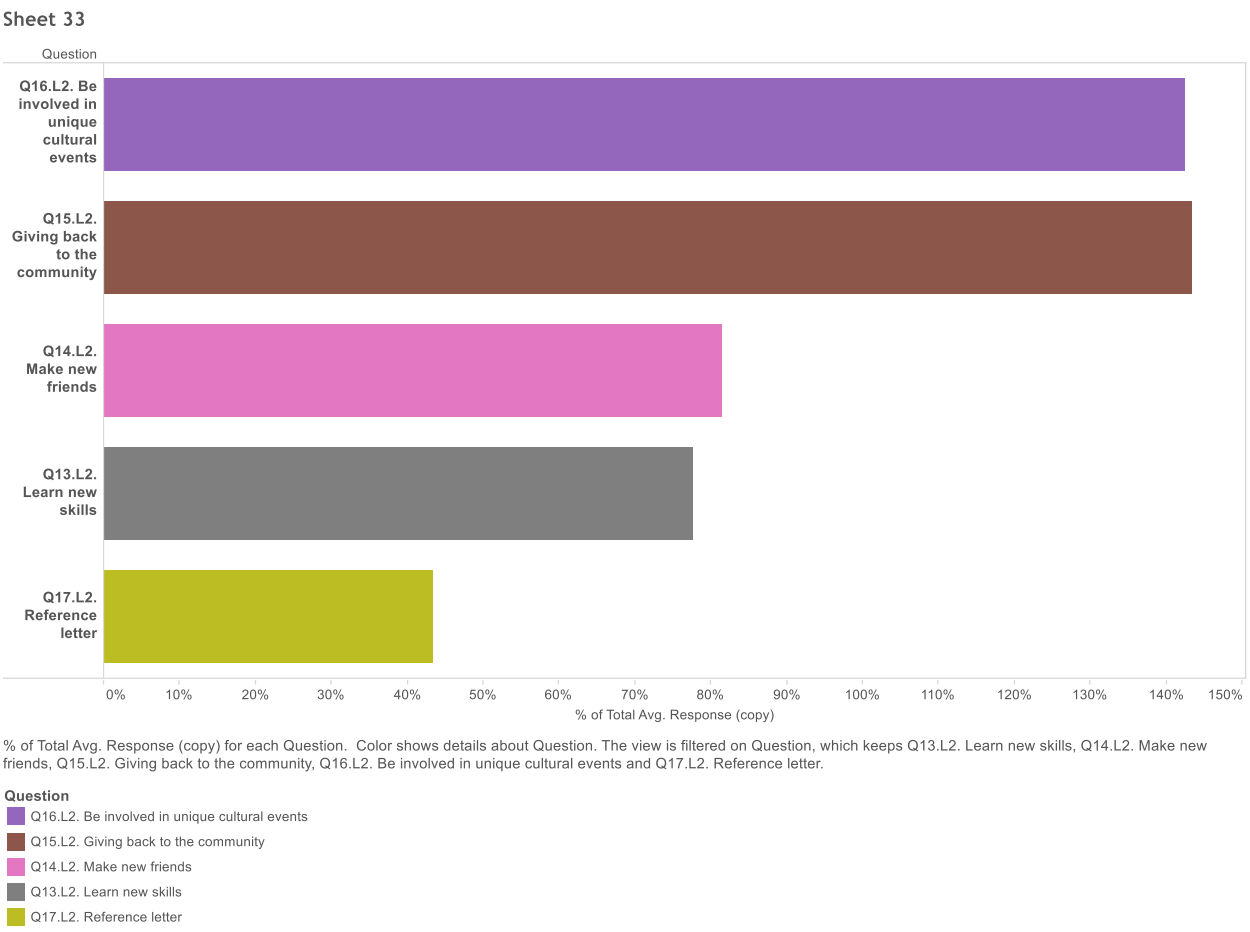
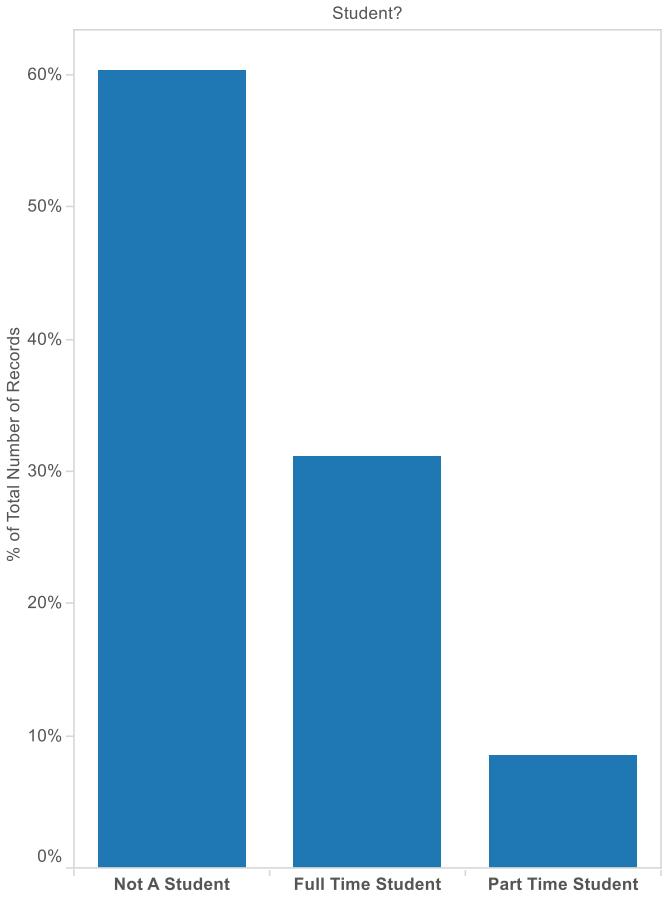


Figure 35. Survey Results: Why are you doing this?

Being involved in cultural events and giving back are the main reasons respondents chose as to why they volunteer.

Sheet 34



% of Total Number of Records for each Student?. The view is filtered on Student?, which keeps Not A Student, Full Time Student and Part Time Student. Percents are based on the whole table .

Figure 36. Survey Respondent Breakdown.

Almost half the respondents are students.

Conclusions and Recommendations

Profile of Volunteers

- Students make up the largest proportion of volunteers at almost 60%
- The average age of a volunteer is 25 with a great many older people volunteering
- Most volunteers are female by almost 3 to 1

Predictors of Hours Worked and Tenure

- There's no difference in tenure or hours worked between men and women
- There is however a significant relationship between age and tenure and age and hours worked (older volunteers are more likely to return and work longer hours)
- There is also a significant relationship between being a student and tenure and hours worked (students work less hours and have less tenure)
- Distance is correlated with hours worked but not tenure (those living < 2 km from Harbourfront put in more hours
- Those volunteers with more than one language work more hours and have greater tenure

Decision Tree and Segmentation

- Four rather distinct segments were identified in the volunteer population:
  1. Young Students Female
  2. Young Males, mostly not students
  3. Old Males Not Students
  4. Middle Age Females – about 41% students

Volunteer Survey

- Overall the volunteers seem very satisfied with the volunteer program
- Their motivation is primarily being part of cultural events, giving back to the community and making new friends

Recommendations

Harbourfront Centre would like to increase the number of returning volunteers as this would help reduce on boarding and training costs.

It would therefore be reasonable to suggest that they target older volunteers who live closer to the Harbourfront Centre. Of course there could be a reason that so many students are accepted and this warrants further discussion.

In addition Harbourfront understandably wants to increase volunteer hours. Fewer volunteers who work more hours are more cost effective, require less scheduling work, and require less training.

Of course it may be impractical to target older non-students who speak another language and live very close to Harbourfront. Perhaps this may even be construed as discrimination. However there are things that Harbourfront could do to fairly encourage more elderly urban people to volunteer. For example, Harbourfront could:

1. Email or send flyers to local retirement homes about the volunteer program
2. Include information about the volunteer program when seniors purchase tickets
3. Organize events for adults and seniors at Harbourfront and make information about the program known

By the same token programs that encourage students to volunteer could be dialed back. University of Toronto makes up the largest contingent of volunteers. Perhaps there's something that is happening that is encouraging students to apply. Maybe students get credit from schools for volunteering, so this practice could be discouraged.

It seems there are a lot of students who are putting in few hours. Perhaps they are doing the bare minimum for course requirements and nothing more. Harbourfront could work on ways to make the experience even more fun for these students so they would put in more hours. It's likely that there's some bias in the surveys and many are discouraged from entering negative answers in the survey. Perhaps a focus group led by a third party would help elicit their true feelings. This could be done in an anonymous way in a completely safe environment. This would provide Harbourfront with some additional information that may prove helpful.

Now that we know a little more about the overall volunteer profile and the general segments to which they belong would could now target those segments differently. We need to know what motivates them. Our second largest segment, middle aged females are about 41% students. The average age is about 38. So these women may be motivated by different things than what would motivate a student. To keep them coming back each year Harbourfront would need to motivate them in a different way than they would a student. Harbourfront currently treats each volunteer in the same way. We know that young students are more likely to be motivated to come back to get credit, however when they are no longer students, how do we keep them coming back? We could keep track of them, send them information about the volunteer program after they are gone, or perhaps encourage them with events that would appeal to them. In the same way, older volunteers are perhaps looking for a way to meet people and learn more about art and culture so there are marketing campaigns that could appeal to them. It could be as simple as sending out a reminder about the program or a card saying "we miss you".

Proposed Next Steps

The following are ideas for further analysis:

1. Develop a better understanding about Harbourfront and their volunteer hiring practices to better understand any constraints or sensitive issues.
2. Further analyze the survey questions and the best way to visualize the responses
3. Explore further analysis to hone in on patterns revealed in this study.
4. Explore the motivation of the volunteers further to best encourage them to return.
5. Try some techniques described to encourage older people to volunteer and monitor the results.

References

Canadians' Arts, Culture and Heritage Activities in 2010, http://www.hillstrategies.com/content/canadians%E2%80%99-arts-culture-and-heritage-activities-2010

Cody, Ron. 2007. Learning SAS by Example a Programmer's Guide. 2013. Cary, NC: SAS Institute Inc.

Statistics Canada, Canada Survey of Giving, Volunteering and Participating, 2010. http://www.statcan.gc.ca/pub/11-008-x/2012001/t/11669/tbl04-eng.htm

Building Caring Communities: The Contributions of Immigrant Volunteers, 2006. http://www.volunteerconnections.net/buildingCaringCommunities.pdf

Perceptions, Attitudes, and Motivations of New Canadians Regarding Volunteerism, 2007. http://volunteeralberta.ab.ca/wp-content/uploads/2011/09/Perceptions-Attitudes-and-Motivations-of-new-Canadians-regarding-volunteerism-CVI-2007.pdf

Making Connections among Canadian Immigrants, 2001. http://volunteer.ca/content/making-connections-social-and-civic-engagement-among-canadian-immigrants

Getting Started with Tableau, 2016 http://www.tableau.com/learn/training#getting-started

Appendix

Here's a list of files included in the submission

SurveryData.xlsx

Includes all the survey data used. There's a tab called "Data Sheet for Tableau" that has the data transformed in the de-normalized format that Tableau requires to analyze survey data.

VolunteersOverview.xls

The original volunteer dataset.

VolunteersOverview5.xls

Reduced data set with only the columns ultimately used remaining.

HFC.twbx

Tableau packaged workbook. All Tableau work is in here.