Popularity and Attendance of Aquatics Recreation Courses*

-Luka Tosic

September 26, 2024

This paper analyzes the relationship between course size and attendance for aquatic courses offered by the City of Toronto's Parks Forestry and Recreation (PFR) commission. The top finding shows that mean student attendance has a directly linear relationship with course registration size among all sign-up based aquatics courses, with deviation in special cases such as swim meets. Further analysis would look at when attendance begins to drop across courses, and where in the city waitlists are higher versus where more courses are being offered. Finally, a comparison to a 2023 realization of this data would allow for analysis of how the Aquatics Department of PFR responded to course demand over the eight years.

```
#init chunk
#| include: false
#| warning: false
#| message: false
#| echo: false

library(tidyverse)
```

```
-- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   3.5.1     v tibble    3.2.1
v lubridate 1.9.3     v tidyr     1.3.1
v purrr     1.0.2
-- Conflicts -------------------------------------------- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
```

---

*Code and data are available at: https://github.com/LukaTosic09/Recreation_Programs_EDA

```
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to beco
```

```
library(dplyr)
library(knitr)
```

1 Introduction

For the City of Toronto's Parks Forestry and Recreation (PFR) Commission's swim instructors and for patrons, course sizes are a constant point of debate. I am a swim instructor for PFR and I am interested in analyzing the popularity and student retention across different courses; as well as analyzing course availability and popularity across the city.

2 Data

The data is taken from (opendatatoronto), it is the Recreational Courses Historical Data dataset. It was cleaned using BaseR, (dplyr) and (tidyverse) packages and visualizations are done with (ggplot). Note that this data is from 2015 and does not relfect changes in course offerings, or any other changes that the aquatics department if PFR has gone through since that time.

Relevant features of the data include categorization by ward and district, classification as drop in or sign up (regular), number of participants registered, the number of classes, the number of participants registered and the number of total visits to a course.

We can't see attendance for each class, but we have the total attendance given by Visits, and the Number_of_Classes to give us a mean visits per week. We can plot the number of students registered vs mean visits over a session. Since the number of students attending each class can't exceed the number registered, if we make a scatter plot and fit a linear model using ggplot, we can expect a slope of at most 1.

First we clean the data and create a bar graph of a sample of courses with over 100 instances.

Count of Total Aquatics Courses
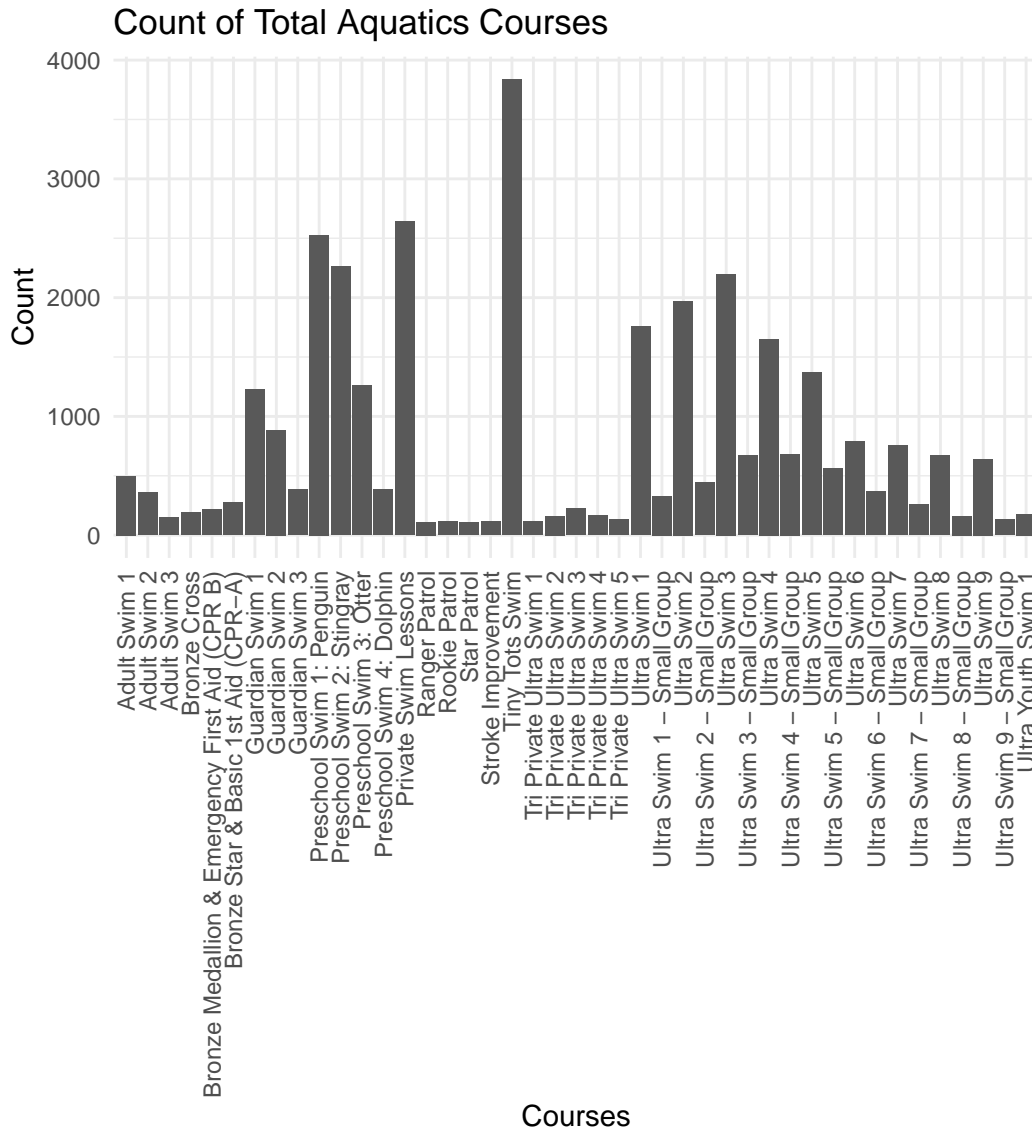
Then we can plot the same count of classes, but now seperated by facility district. Note this data includes only courses with more than 50 instances, for readability.
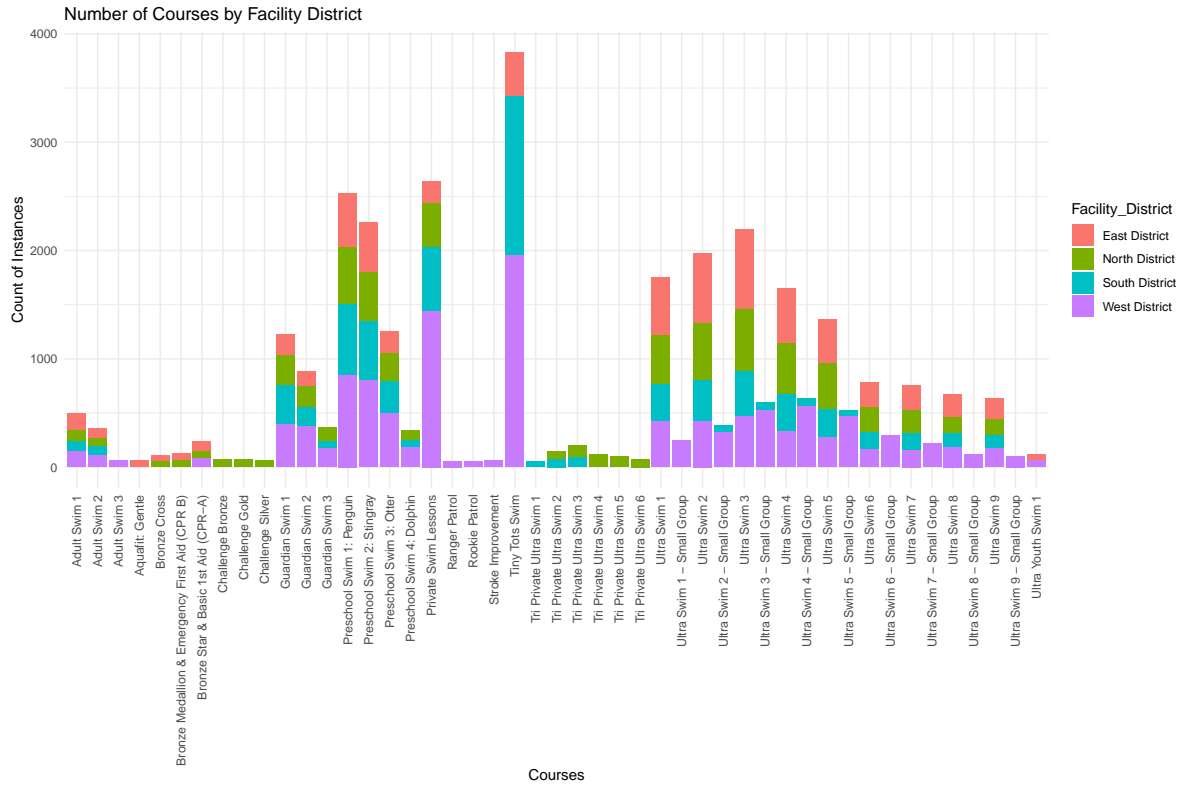
3 Discussion

Tiny Tots Swim is by far the most offered in the city, this makes sense as the classes are only 15 minutes as opposed to all other classes which are 30 minutes or longer, so more can be offered in a single day; moreover, Tiny Tots can be run at any pool including wading pools which increases the number of facilities that can host it.

We can also see that the small group ultra 1-9 swims are offered anywhere from twice to five times less often than their regular counterparts. The tri private swims are 3-student courses offered most commonly at Douglas Snow Aquatic Centre (5100 Yonge St, North York) in the North District and Matty Eckler Recreation Centre (953 Gerrard St E, Toronto) in the South District.

Though this is a sample of all aquatics courses offered, we can see that they are present throughout the city, but with the South District which is the Downtown area, offering the least courses. This is likely due to the lower density of pools in that district.
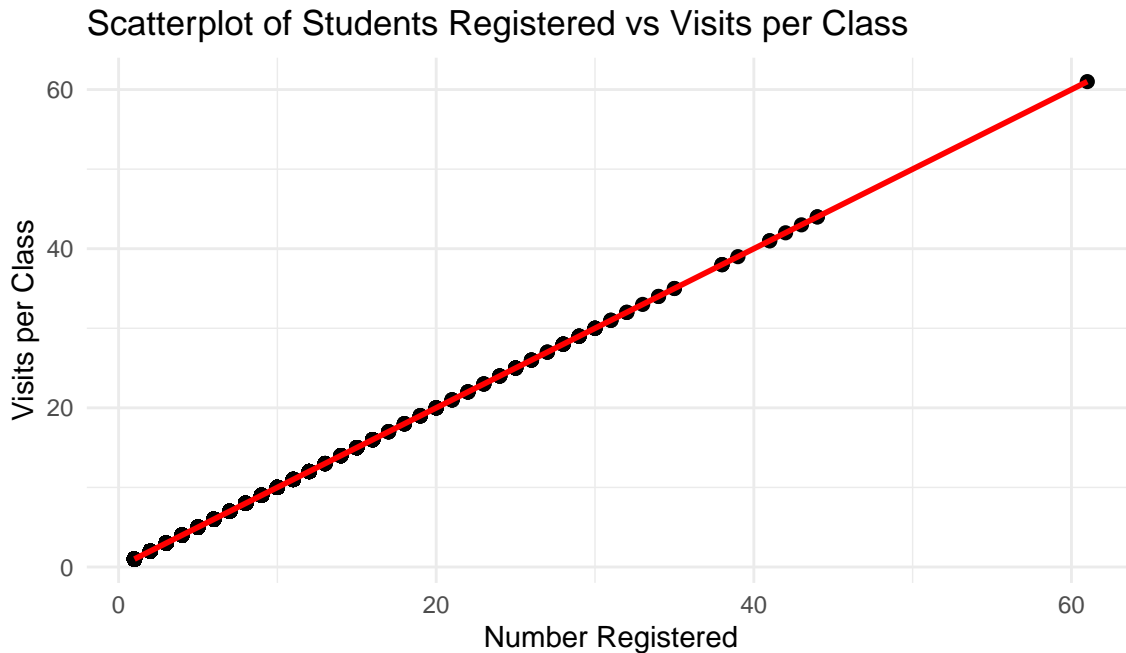
Scatterplot of Students Registered vs Visits per Class

Data constrained only to allow points for which the number of visits per week does not exceed the number of students registered.

Scatterplot of Students Registered vs Visits per Class

4 Discussion

We see as we expected a slope of exactly one for mean visits per week across all instances of a course; meaning that a one unit increase in the amount of people registered for a course would result in a mean increase of one participant per class.

Anecdotally, I would say that students continue to show up regularly if they have showed up for the first 3 classes for a course in a given session. For those that do not show up, they more often than not drop out and are replaced by people on the waitlist or by late applicants to a course by the third week. Any small deviations in the number of participants for a class would be too minuscule to show up in the mean.

The first scatterplot shows a large number of courses in the West District that have significantly higher attendance than the number of people registered. I am not entirely familiar with how the data was collected; however, Etobicoke Olympium Pool (590 Rathburn Road, Etobicoke) in the West District is an Olympic sized pool that hosts virtually all of the City's municipal youth Swimming competitions named "Splash". It is a reasonable assumption to make that those competitions are what is being captured by the data, as the Splash meets do show up in the dataset.

5 Weaknesses and Next Steps

This data can only tell us about average attendance as the Visits variable gives the sum of all instances of patrons attending a program. To analyze further, we would want to look at when patron attendance drops occur, and for what courses/programs. Is there a trend in this? To do this, we would need a vector that stores visits by class. As an instructor with the city, I know this is possible as our attendance sheets contain a box for the total number of participants each week; however the data was logged as total participation over all weeks.

Another next step would be a further analysis of waitlist sizes for different courses by ward and district to analyze where certain courses are lacking in availability. A scatterplot of total waitlisted individuals over all instances of a course as a response, with course instance Count as a predictor, split among wards or districts would allow us to see where the demand for courses is and what courses are in demand, and where and how this demand is being supplied. It would be important to note that we expect that more popular courses would have a higher waitlist only if there isn't sufficient demand, thus an assumption of linearity would be a large one to have to make.

A final step would be to compare this data and its analyses with a realization of this dataset from closer to the present, ideally 2023, to see how trends have changed in the time that has passed.