



Final Project Process Book

Data Visualization C Term 2019

Kylie Dickinson, John Dyer, Josh Galang, Demi Karavoussianis

Overview and Motivation

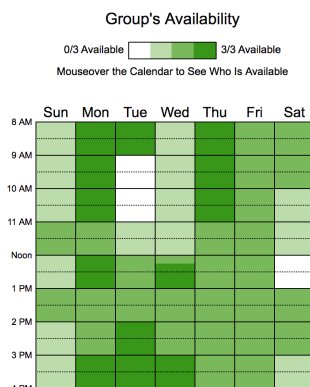
For many students and faculty, the recreation center (gym) at Worcester Polytechnic Institute (WPI) is a daily distress and work zone. Due to students having similar course and activity schedules, the gym becomes very busy at specific hours of the day. These peak hours make exercising more difficult and time consuming since gym supplies are limited and must be shared. In order to avoid these busy hours, our project team wanted to determine at what hours the gym is the least crowded. The goal of this project was to obtain the gym entrance data, analyze it, and create visuals that will help show when the best time is to go to the gym for each day of the week.

Related Work

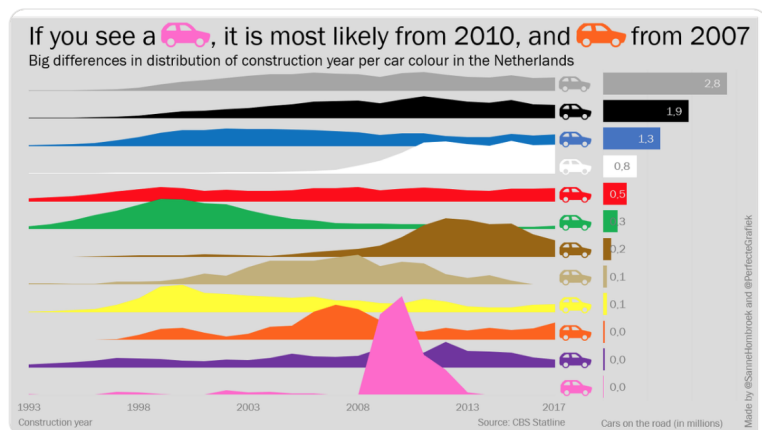
We were inspired to create a heatmap, especially after frequently filling out When2Meet for classes and extracurricular activities. When2Meet is an interactive group availability heatmap, where each user highlights the time they are available, and a heatmap is automatically generated based on group availability, with the “best time” to meet in the darkest shade of green. <https://when2meet.com>

After reviewing the data, our group remembered a visual reflected on in class which used area line graphs to display information. We revisited this graph at the following link, <https://twitter.com/PerfecteGrafiek/status/1095074427895799808>. Although this area chart does not represent the best data visualization, the concept inspired us to create a series of area charts for the average entries per hour, with a linked tooltip functionality. We purposefully made sure that the issues with the original graph, such as color, spacing, and overlapping, were fixed and improved for our visualization.

Other inspirations for the visuals we created came from creativity and knowledge learned from the textbook.



When2Meet



1:37 PM - 11 Feb 2019
Reflection Visual

Questions

Our project proposal lead our group to ask questions about the WPI recreation center traffic patterns and usage statistics. These questions were:

- What do traffic patterns look like at the WPI recreation center?
- Is it possible to determine the least crowded (best) time to go to the gym?
- Who uses the gym most?
- Is there a time of day where one gender out populates the other?

Data

Data was collected from Meredith Merchant, the Assistant Director of Facilities & Recreation. The recreation center uses a software that collects information on all users who access the gym. Students, faculty, guests, and paying users must access the gym by using their WPI ID. The WPI ID, along with information linked to that ID (gender, student, etc.), is recorded into the software for the time at which the user accesses the gym. Meredith Merchant had access to this information and provided our group with data sets from the month of February. The month of February was chosen due to its fairly normal gym hours (no special events) and it being in the middle of a term, not finals week where data could be skewed.

Our team was provided with three datasets all containing recreation center entrance information from February 1st through February 24th. The first contained hourly totals for 6a to midnight of each day of the week. The next was the same data only broken down by annual membership (student, paid membership, faculty). The final data set contained information on gender and membership type for each hour of the day range. Data that was in total counts were divided by 24 (the number of days in which data was collected) in order to find the average usage for that time frame. These three datasets were then reformatted from excel sheets to csv files in order create the data visualizations.

Exploratory Data Analysis

Data was initially viewed in an excel worksheet where insights about the sizes of the datasets and variables of each dataset were made. The first data set was the simplest with eighteen different time ranges for the seven days of the week. This reminded the group of a typical when2meet array and other heatmaps reviewed in class. The next dataset had information about the different memberships, and since we didn't care about population breakdowns at individual time points, and rather the average usage of each membership type, we knew we would need to show the data as percentages. Moving onto the third dataset, the group observed that gender was recorded similarly to the way in which memberships were recorded. The group decided it would be important to show this data as comparisons of each population (male vs female) at each time range in an average day. After reviewing all the datasets in excel, the team proceeded to brainstorm potential visualization designs.

Design Evolution

Our project team brainstormed the following list of possible visualizations that could be created with the data provided in the datasets.

Final Project Group Brainstorm

Gym Users per Hour Information

- Area Line Chart*
 - days of the week vs time
 - Ex. <https://bl.ocks.org/d3noob/119a138ef9bd1d8f0a8d57ea72355252>
- Parallel Coordinate Chart
 - Each line is a day of the week, the y-axis is the population at a given hour, and the x-axis contains lines for each hour of the day
- Heatmap*
 - Showing populations by gradients for hours of the day for each day of the week
 - When2Meet - <https://when2meet.com>

Membership Types

- Clustered Bar Chart
 - Each cluster contains information about each group, each cluster is at a different time point
- *Pie Chart
 - On average, percent of users in each category would be the slices of the pie

Gender Information

- *Diverging Histogram
 - Two horizontal histograms mirrored on an axis with one histogram containing male information and the other containing female
- Pie Charts
 - Two categories could be shown in 18 graphs where each graph represents an average hour of the day and the information contains the percent of males vs female users
- Clustered Bar Chart
 - Each cluster has a bar for male and female, each cluster is at a different hour for the average day

From the outline above, our team chose to create the starred visuals. For the first visual, we created an area line chart that contains area line charts for each day of the week, stacked on top of each other with time ranges aligned in order to compare busier to slower times at the gym on average for each day of the week. Our group decided to use multiple area charts instead of the parallel coordinate chart for easier readability and comparisons. For an average user, stacked area charts are easier to visualize as each graph describes gym populations for a given day of the week in order from Sunday to Saturday. The parallel coordinate graph would make it more difficult to identify gym populations for individual days.

The next visual our team wanted to create was a heatmap. The data reminded our group of a when2meet, so we created a similar visual. This made sense since we are trying to display which hours

are busy vs slow at the gym and the heatmap makes this easy to identify the lighter colors which would represent slower hours versus darker colors that represent busy hours.

The third visual our group decided on was a pie chart. The pie chart is a common and simple graph, and since we had information on usage of different memberships, we decided that determining who was actually using the gym would be interesting. The pie chart simply and easily displays the average percent of which groups use the gym most versus the least.

The final visual our group decided to pursue was the diverging bar graph. We wanted to be able to determine usage by both male and females and then compare that data. This could have been done with multiple pie charts; however, that would take a lot of space and make it difficult to compare hours to each other. It also could have been done using clustered bar charts, but our group decided that the diverging bar graph was more aesthetically pleasing and interesting to observe. The diverging bar graph also allowed for each group to be compared separately.

Implementation

Multiple Area Chart Visual

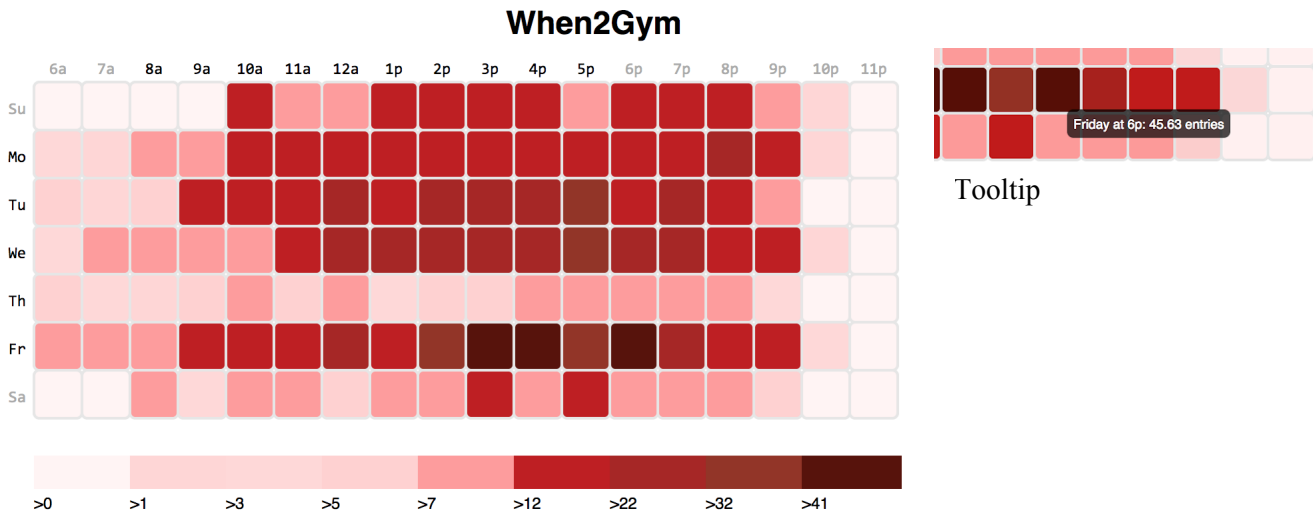
The multiple area charts visual was created in order to compare day to day variations in populations at the gym. It was intended to allow for a user to look at the chart and determine the difference between slow and busy hours at the gym. The graph uses a hover tooltip to provide the user with more specific information for each hour of the day. The tooltip displays the average amount of gym users for the chosen time of day for each day of the week. Hovering the mouse not only displays a tooltip but highlights the hour of the day on each area chart. This should allow the user to determine how busy the gym is at a given time of day for each day of the week by providing more detailed information than can be observed by just viewing the visual.



Above: Tooltip
Left: Visual

Heatmap Visual

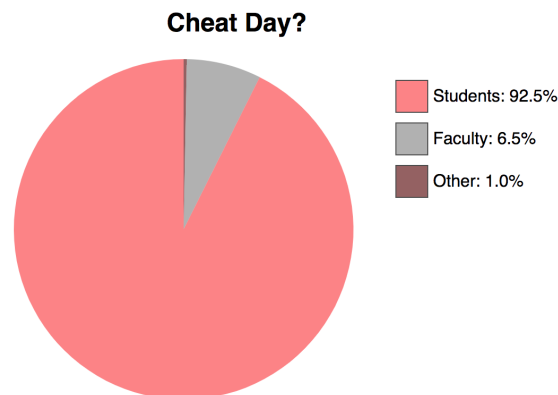
The heatmap was intended to allow users to observe times of high and low populations at the gym. The heat map compares busy hours to slow hours through gradients of dark to light colors. The heat map uses a hover tooltip that allows for users to mouse over any given box and be provided with the exact number of average users at the gym for that given time point.



Visual

Pie Chart Visual

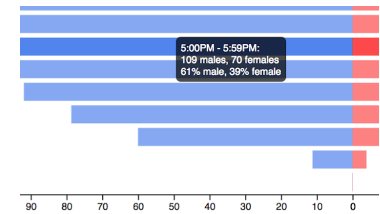
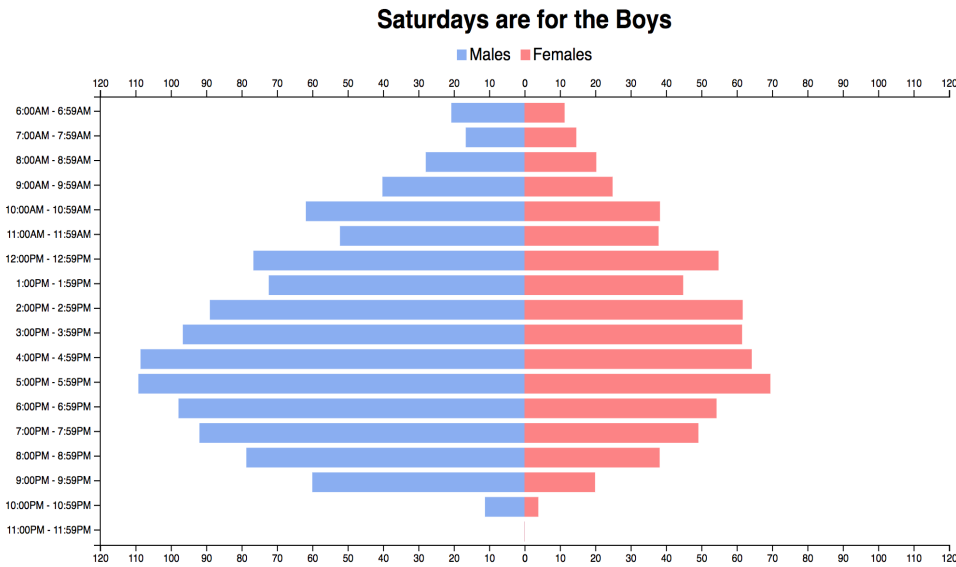
The pie chart visual was simply intended to display the average usage of each membership type at any given time for any given day. Our group wanted this to be simple since it was not a question our team was initially trying to research, however, we were given the information and wanted to make use of it. Originally the graph was labeled at each slice, but those labels were switched to a legend for readability and visual purposes.



Diverging Bar Graph Visual

The intent of the diverging bar graph was to be able to observe the daily averages of male and female populations per hour at the WPI recreation center. The center line divides the two horizontal

histograms which allows for the values of each population to be compared. The visualization has a hovering tooltip which provides the time, the total number of males and females, and percentage of each. Since there are multiple bars and values on each axis, the tooltip helps observe specific statistics at each time point.



Tooltip

Visual

Evaluation

From the visuals, our team determined that, though there are similar traffic patterns, it can depend on the day for what time range is the best (least crowded) to go to the gym. The visuals answered the questions we previously asked. The first and second visualizations answer at which time points the gym is least crowded and also at which times the gym is the busiest which describes the traffic patterns of the recreation center. The pie chart describes what type of memberships most frequently use the gym. The last visualization, the diverging bar graph, displays at which hours of an average day the gym is most populated with males versus females.

The visualizations work to the extent we had hoped by efficiently portraying information that answers our research questions. Further work on the visualization would be useful in order to make the visual more accurate and successful. With more time, our group would have liked to have used a larger data set that spanned not just the month of February but the entire school year. With that information, we would have determined 'special event' days (open house, accepted student's day, career fair, etc.) that would need to be excused, and even processed for their own analyses. Other visualizations our group would have liked to implement were monthly breakdowns as well to describe at which time of year the gym is the busiest, such as after New Years/ Christmas break or the beginning of A-Term, versus the least busy such as finals week. We would have also liked to improve the implemented visuals. Our group would have liked to add a tooltip that scrolled over the legend of the heat map and for each population range, 'light up' or add outlines to the squares that represent those ranges so that users could easily

determine at which times the fewest people were at the gym since the gradient can sometimes flow together. Lastly, another idea our group had was to link the visuals to show how they relate, such as adding a tooltip that displayed percentages of male and female users for each membership of the pie chart or adding a tooltip membership/gender information to the diverged bar graph tooltip.

Overall, our visualization answered all of our initial research questions as well as others. The visualizations are user friendly which was a goal of our website so that students could actually use the tool to find a better time to work out. The project created the first steps to a potentially greater project that could be completed if more data was collected from the recreation center staff, surveys, etc. This would allow for other visualizations to stem from the basic visuals our group outlined.