



Swimming Performance Capstone Project

Names: Megan Christy, Julia Miraglia, Omkar Sakhawalkar, Shwetha Venkatesh

Background

Motivation:

The CMU Varsity Women's Swim and Dive Team has drastically improved over time and we are aiming to analyze this improvement.

About Swimming:

4 strokes in swimming: butterfly, backstroke, breaststroke, freestyle
Varying distances: 50 yards all the way through 1650 yards for some strokes
Swimmers tend to have specialty strokes and distances and tend to compete in the same events.

Client:

Coach Kinney: CMU Men's and Women's Swimming & Diving Coach

Goals:

Analyze improvement in performance for swimmers using historical data (2002-current)
Create visualizations from data to highlight trends in women's team since 2002 season to present to recruits

Data Sources / Data Processing

Original Data:
PDF's and data posted directly on CMU Athletics Website

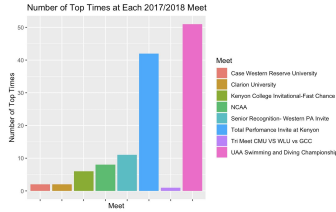
Changed Meet column to have consistent meet names across seasons

| Date | Meet | Date | Meet |
|------------|--------------------------------------|------------|---------------------------------------|
| <Date> | <Meet> | <Date> | <Meet> |
| 2005-02-10 | UAA Championship 2005 | 2005-02-10 | UAA Swimming and Diving Championships |
| 2005-02-10 | UAA Championship 2005 | 2005-02-10 | UAA Swimming and Diving Championships |
| 2005-02-10 | UAA Championship 2005 | 2005-02-10 | UAA Swimming and Diving Championships |
| 2005-02-10 | UAA Championship 2005 | 2005-02-10 | UAA Swimming and Diving Championships |
| 2005-01-15 | Carnegie Mellon vs Westminster Colle | 2005-01-15 | Westminster |
| 2005-02-10 | UAA Championship 2005 | 2005-02-10 | UAA Swimming and Diving Championships |
| 2005-02-10 | UAA Championship 2005 | 2005-02-10 | UAA Swimming and Diving Championships |
| 2005-01-08 | Carnegie Mellon vs Case Western | 2005-01-08 | Case Western Reserve University |
| 2005-02-10 | UAA Championship 2005 | 2005-02-10 | UAA Swimming and Diving Championships |
| 2005-02-10 | UAA Championship 2005 | 2005-02-10 | UAA Swimming and Diving Championships |

Data entered manually into Excel then transferred to R

| Season | Event | Time | Name | Age/Yr | Date |
|-----------|---------------|--------|------------------|----------|------------|
| <Season> | <Event> | <Time> | <Name> | <Age/Yr> | <Date> |
| 2002/2003 | Women 50 Free | 25.10 | Julie Tibor | JR | 2002-11-23 |
| 2002/2003 | Women 50 Free | 25.47 | Me-Lan Ho-Walker | SR | 2003-02-12 |
| 2002/2003 | Women 50 Free | 25.65 | Conny Goldberg | JR | 2003-02-12 |
| 2002/2003 | Women 50 Free | 25.88 | Audra Kormanik | SR | 2003-02-12 |
| 2002/2003 | Women 50 Free | 26.49 | Jessica Gibbons | SO | 2002-11-02 |
| 2002/2003 | Women 50 Free | 26.63 | Shelli Coats | SO | 2002-11-23 |
| 2002/2003 | Women 50 Free | 26.87 | Kate Kiche | SO | 2003-02-12 |
| 2002/2003 | Women 50 Free | 26.76 | Jany Yan Olfen | JR | 2002-12-06 |
| 2002/2003 | Women 50 Free | 26.99 | Jenny Wong | SO | 2003-01-11 |
| 2002/2003 | Women 50 Free | 27.85 | Allison Theus | FR | 2002-10-05 |

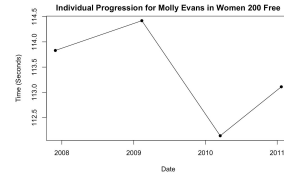
Exploratory Data Analysis



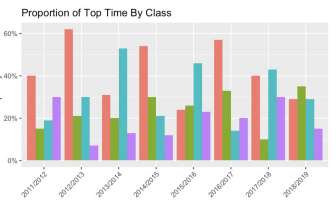
We notice that top times are recorded most frequently at the Total Performance Invite, and the UAA championships. These are the meets that the swimmers are specifically training for and the ones they taper for (planned decrease in training volume to maximize performance).



Plotting time against year for each event, there is a clear negative relationship. This means that the team has gotten faster.



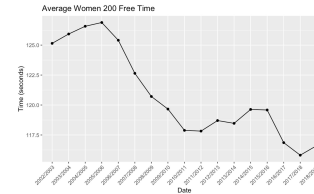
This follows the progression of a single swimmer over their career in the event '200 Free', from 2008-2011. The function can provide visualizations subsetted by swimmer, allowing our client to follow his athletes progress closely.



This graph displays the proportions of top times by Age/Class since the 2011/12 season. Freshmen have performed well relative to upperclassmen for all years, especially earlier seasons.

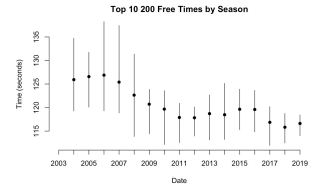
Results

We created a function/series of functions that allows Coach Kinney to visualize trends. He can subset based on event, swimmer, season, and meet. The function outputs a graph filtered on the desired subset over the season, or over a desired period of time, using the data processed from 2002 to 2018. It also outputs summary statistics for each event.



While there is some variation, there is a clear negative linear trend in the average 200 freestyle time for women. All events exhibit similar trends, indicating that the women's team is getting faster

The vertical lines represent all 200 free times for each season. It appears that in general, the variation in the top times has gone down, indicating that the women are performing at more similar speeds.



```
[1] "2002/2003"
  Min. 1st Qu. Median Mean 3rd Qu. Max.
  119.0 121.2 124.3 125.1 127.0 136.8
[11] "2003/2004"
  Min. 1st Qu. Median Mean 3rd Qu. Max.
  119.2 122.3 125.7 125.9 129.0 134.7
[11] "2004/2005"
  Min. 1st Qu. Median Mean 3rd Qu. Max.
  120.1 122.6 127.0 126.6 131.1 131.7
```

The function can produce summary statistics of the filtered data based on what is inputted. This example is of the top times for Women's 50 Free by year.

Next Steps

- Convert functions to user-friendly R shiny app
- Duplicate project for men's team
- Come up with a better way for data collection
- Implement across all meet results, not just top times