Exercise: Variability and Uncertainty

This exercise uses:

- the `{ggplot2}` library and a custom function
- Your knowledge from past in-class exercises, videos, homework, etc. and corresponding modules from the course site.

Overview

This exercises provides some practice creating visualizations that allow you to communicate uncertainty or variability in data. Whereas `geom_bar()` or `geom_col()` *may* be useful for communicating summary statistics like the mean or median, they fail to communicate information about uncertainty or variability in data. `geom_histogram()`, `geom_density()`, and variants thereof visualize the distribution but this information may not always be needed. In some instances, you need to communicate dispersion statistics, confidence intervals, inter-quartile ranges, or other information about a distribution.

A research paper on misinterpretation of error bars by *Krzywinski & Altman (2013)* can be found **here** if you are uncertain about comparing distributions by their standard errors and/or confidence intervals.

Data Set

Use the `SWIM` data set or your project data set if appropriate. You can access the `SWIM` data here.

Custom Function

The following function will take a vector, compute its mean and standard error in order to estimate the confidence interval. What is returned will be a data frame holding three column variables, `y`, `ymin`, and `ymax` which represent the mean and the upper and lower limits of the confidence interval. You should use the code to create a function in your script or R Markdown file so that you can use it in your answer.

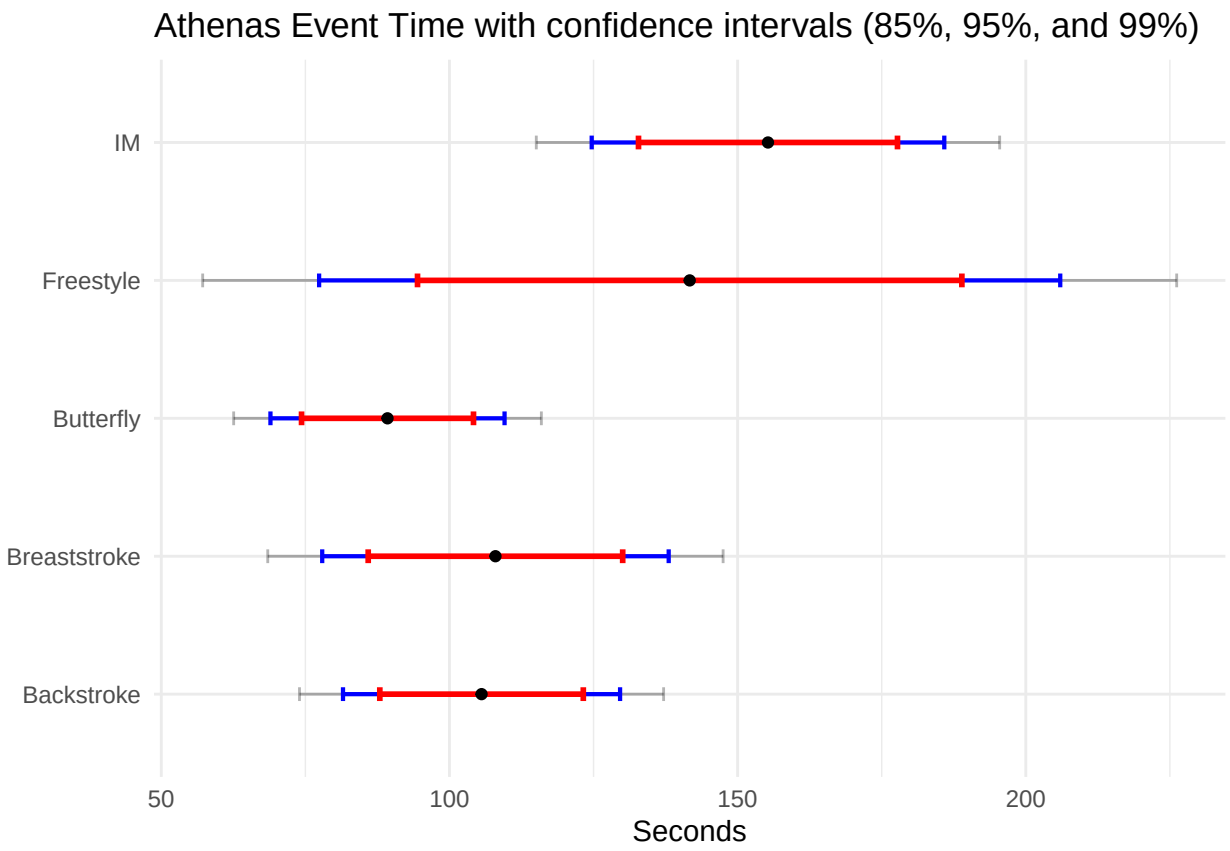
```
mean_ci <- function(x, level = 0.95) {
  x = na.omit(x)
  m = mean(x)
  se = sd(x) / sqrt(length(x))
  ci =  qnorm(1 - (1 - level) / 2) * se

  return(
    data.frame(y = m, ymin = m - ci, ymax = m + ci)
  )
}
```

NOTE: Your custom function will iterate your grouping variable. As covered in other content, use the `~` to perform a lambda function along with `.x` to represent the variable.

Example `~ my_function(.x)`

Problem: Plot Replication

Replicate this Plot as best as you are able.


Athenas Event Time with confidence intervals (85%, 95%, and 99%)
IM
Freestyle
Butterfly
Breaststroke
Backstroke
50
100
150
200
Seconds