# Mission Analytics Data Exercise

Andrew Han

# Exercise Questions

Using the data provided, create a series of visualizations (e.g., charts, graphs, etc.) that answer the following questions regarding the results of the 2006 Pike's Peak 10k Race:

1. What are the mean, median, mode, and range of the race results for all racers by gender?

2. Analyze the difference between gun and net time race results.

3. How much time separates Chris Doe from the top 10 percentile of racers of the same division?

4. Compare the race results of each division.

# Tools and Resources

Analysis done using Python (3 hours)

Google Slides (0.5 hours)

Code, Exercise, Graphs available on Github: https://github.com/DrewHanSolo/MarathonAnalysis

# Presuppositions and Disclaimers

In the interest of time...

- Runners that have misentered data are ignored. No attempts made to correct data
- Time durations are presented in seconds
- Runners with a Net Time < 500 secs ignored for most of analysis
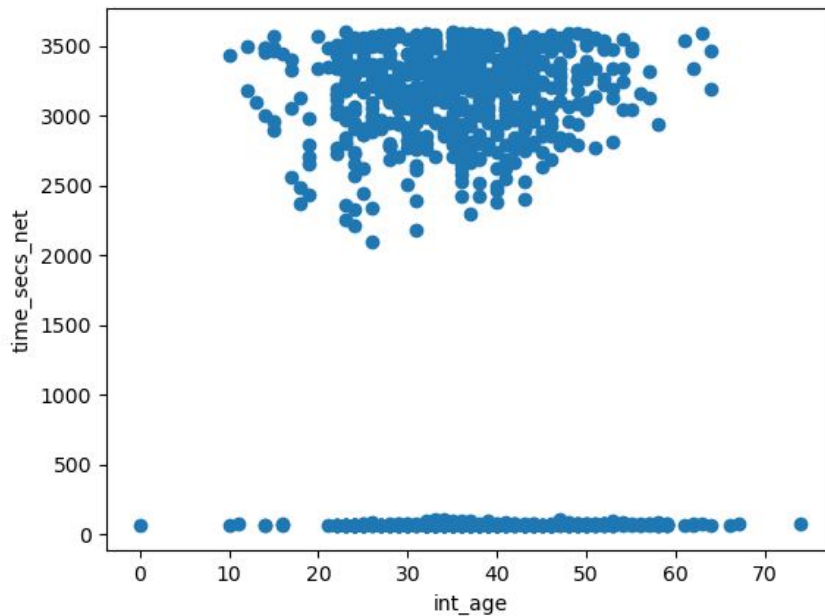- No unit tests performed on analysis code

More analysis information available in Analysis.out and Parser.out

# Raw Data (valid runners)



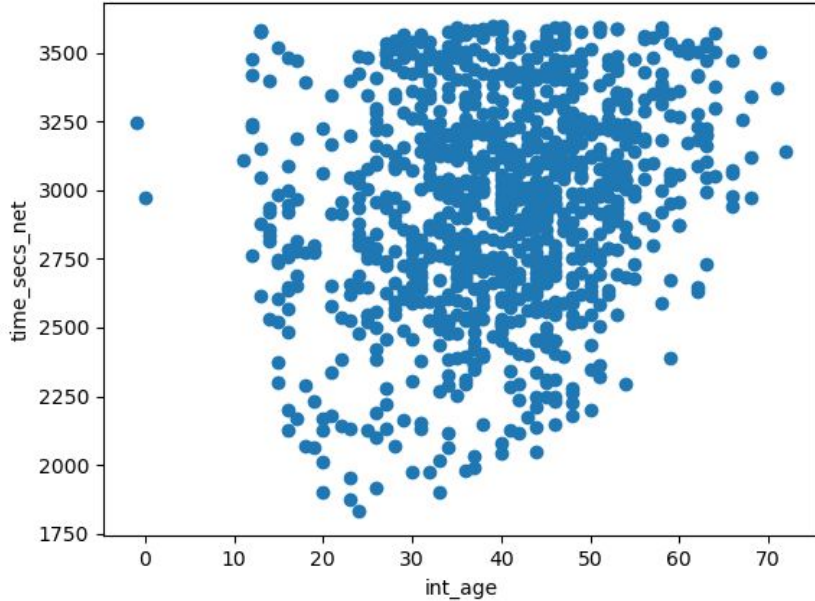Male Runners: scatterplot of int_age vs time_secs_net (1240 runners)
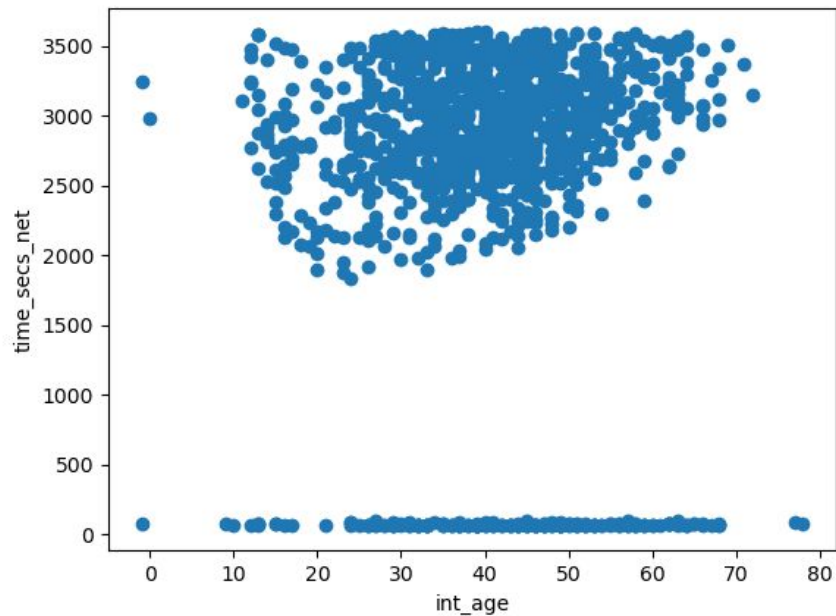
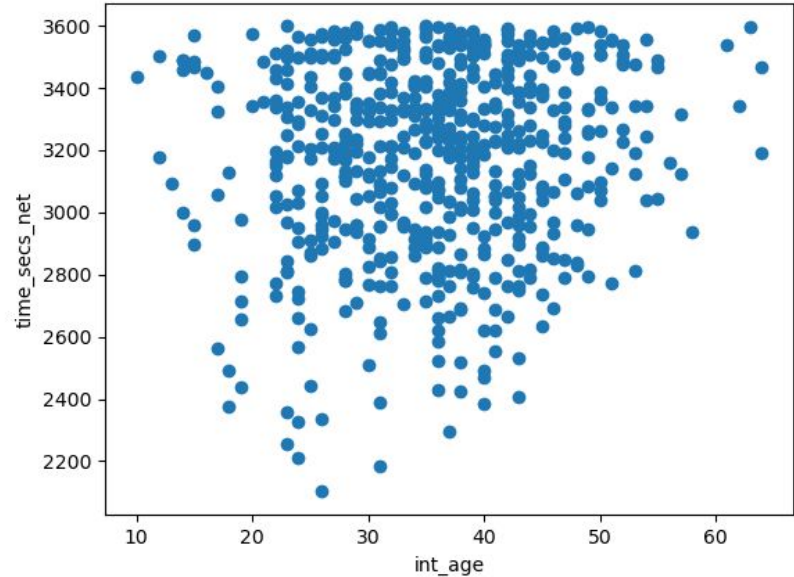Female Runners: scatterplot of int_age vs time_secs_net (946 runners)

# **Raw Data** (valid runners, NetSecs>500)



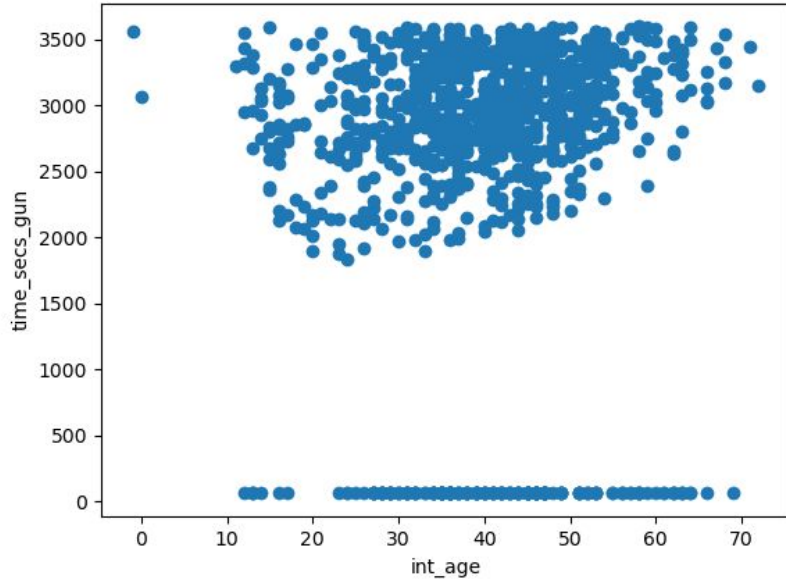Male Runners (time_secs_net > 500): scatterplot of int_age vs time_secs_net (1041 runners)

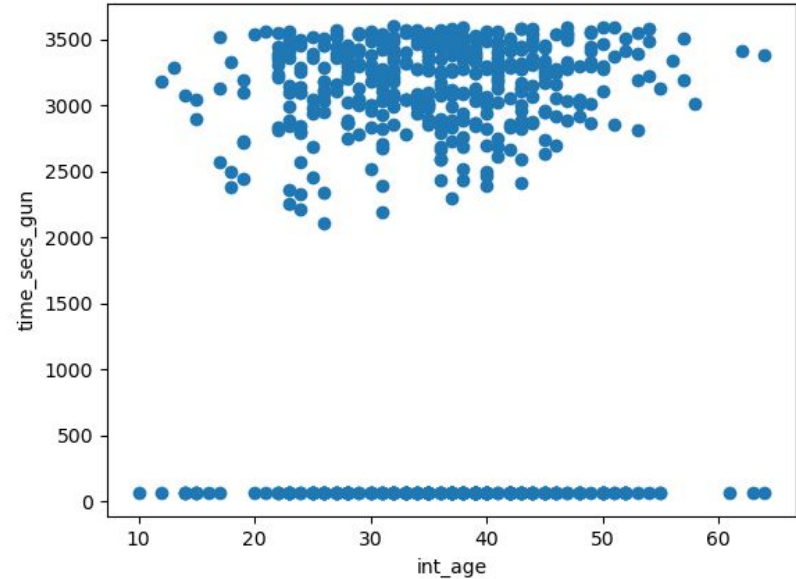Female Runners (time_secs_net > 500): scatterplot of int_age vs time_secs_net (603 runners)

# **Raw Data** (valid runners, NetSecs>500)



Male Runners (time_secs_net > 500): scatterplot of int_age vs time_secs_gun (1041 runners)

Female Runners (time_secs_net > 500): scatterplot of int_age vs time_secs_gun (603 runners)

Even filtering erroneous NetSecs<500 runners, GunSecs<500 still present.

**2. Analyze the difference between gun and net time race results.**

Given the erroneous inputs of NetSecs and GunSecs, the difference between values does not hold much meaning. No filters on this calculation:

Male Runners: mean_time_secs_net = mean_time_secs_gun + 378.154032 secs

Female Runners mean_time_secs_net = mean_time_secs_gun + 576.133192 secs

Takeaway: Females crossed the start line later than males

## 1. What are the mean, median, mode, and range of the race results for all racers by gender?

Ignoring runners who dropped out of the race... (Net Time < 500 secs)

Skipping graphical visualization...numbers are pretty close and graph not needed for comparing 2 entries. See slides 5, 6, 7
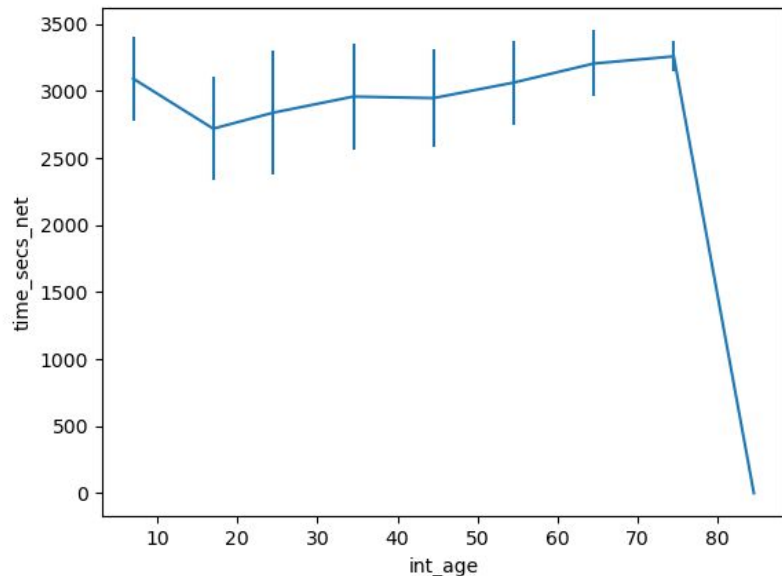
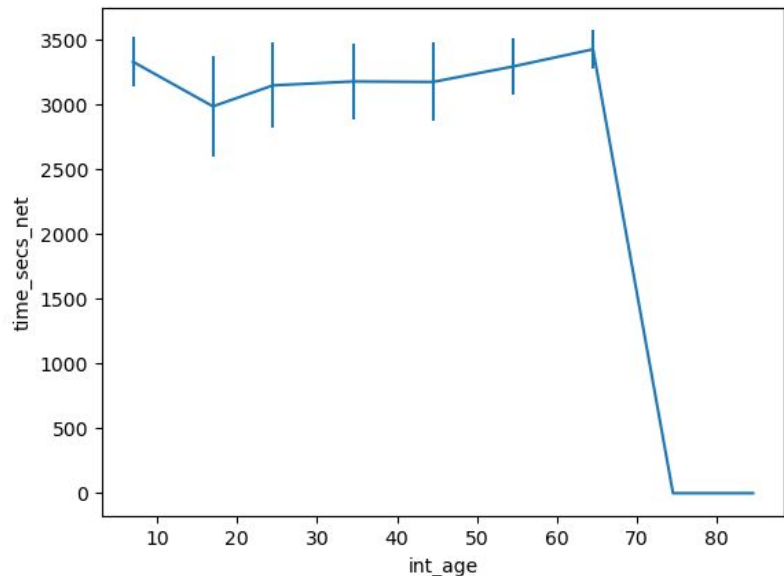| Male Runners (time_secs_net > 500): | | | Female Runners (time_secs_net > 500): | | | Female Differences From Male | |
|---|---|---|---|---|---|---|---|
| Attribute = time_secs_net | | | Attribute = time_secs_net | | | | |
| Mean | 2962.10951 | | Mean | 3176.383085 | | 214.2735745 | |
| Standard Dev | 387.778512 | | Standard Dev | 306.0706909 | | -81.70782108 | |
| Median | 2988 | | Median | 3226 | | 238 | |
| Mode | 2128 | | Mode | 2886 | | 758 | |
| Range = [1830, 3597] | | | Range = [2102, 3599] | | | | |
| Count = 1041 | | | Count = 603 | | | | |

# 4. Compare the race results of each division.

divisions = [[0, 14], [15, 19], [20, 29], [30, 39], [40, 49], [50, 59], [60, 69], [70, 79], [80, 89]]



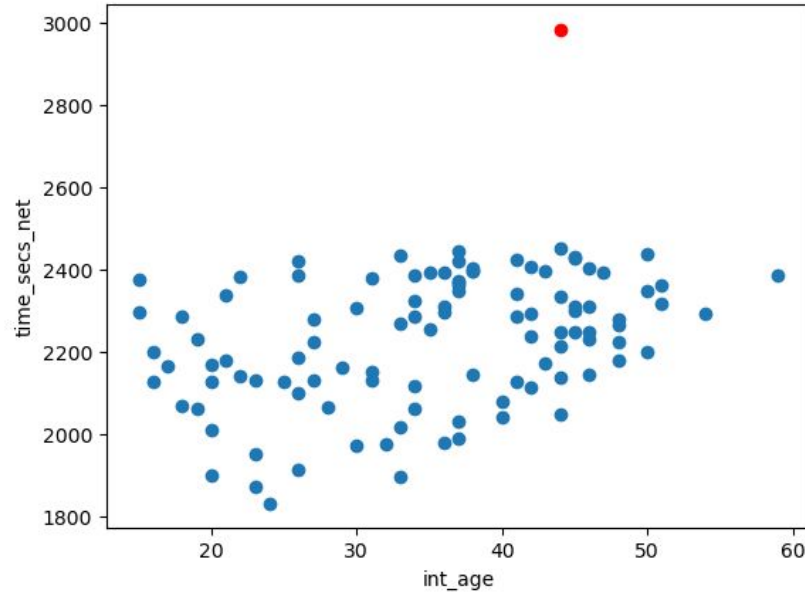Male Runners (time_secs_net > 500): avgs of time_secs_net binned by int_age (1041 runners)



Female Runners (time_secs_net > 500): avgs of time_secs_net binned by int_age (603 runners)

**3. How much time separates Chris Doe from the top 10 percentile of racers of the same division?**



Chris Doe compared to division top10%: scatterplot of int_age vs time_secs_net (105 runners)

Net time difference of Chris Doe from top 10% male runners: 754.466667 secs

# Future Improvements

- Attempt to sanitize misentered data, corroborate entries against each other
- Extend analysis code to layer and color code graphs by sex and age divisions
- Add unit tests to analysis code
- Get a Microsoft Excel license and export raw analysis data to workbook
- Reformat log output