

# 2006 Pike's Peak 10k Race Results: Data Exercise

Jorge Rodriguez, MPH  
July 12, 2020

# 2006 Pike's Peak 10k Race Results: Summary

- **Data Cleaning**
- **Question 1:**
  - *What are the mean, median, mode and range of the race results for all racers by gender?*
- **Question 2:**
  - *Analyze the difference between gun and net time race results.*
- **Question 3:**
  - *How much time separates Chris Doe from the top 10 percentile of racers of the same division?*
- **Question 4:**
  - *Compare the race results of each division.*

# 2006 Pike's Peak10k Race

## Results: Data Cleaning

- Missing Data
  - Adding default values or use mean to fill data
  - Replace missing data
- Data Types
  - Check and change to data type needed to analyze
- Generate Variables
  - Create variables needed for analysis
- Remove/Rename Typos
  - Human Error

### Generating New Variables: Difference Minutes

Gun Minutes	Pace Minutes	Difference Minutes	Places	AgeGroup	New_Divisor
28.8	4.63333	0.0166667	1-49	20-29	3
29.1833	4.7	0.0166667	1-49	20-29	3
29.6333	4.76667	0.0166667	1-49	20-29	3
29.7667	4.8	0	1-49	20-29	3

### Data Types: Pace - Object to a TimeDelta64

Pace	Gender	Division ▲	Total	Net Minutes	Gun Minutes
0 days 00:04:48	M	1	96	29.7667	29.7667
0 days 00:05:14	M	1	45	32.5	32.5
0 days 00:05:26	M	1	132	33.7	33.7
0 days 00:06:04	M	1	50	37.65	37.65

### Net Time - Removing ex. (“#”, ”\*)”)

Hometown	Gun Time	Net Time	Pace	Gender
Kenya .	32:59	32:58#	5:19	F
Kenya .	33:14	33:13#	5:21	F
Gaithersburg MD	34:26	34:25#	5:33	F
Great Falls VA	34:43	34:42#	5:35	F

# 2006 Pike's Peak 10k Race

## Results: Question 1

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

# of Racers: Males = 1265 & Females = 1105

1. The mean, median, and mode for our race results is higher overall for females than males.
2. For both Males and Females the data is positively skewed because the means for both genders is higher than the mean and mode, respectively.
3. Because of the positive skewness, the median is a better indicator of where the central data is for both genders.
4. This is also supported by the number of outliers from both genders, later shown with boxplots.

Mean			
Gender	Net Minutes	Gun Minutes	Pace Minutes
F	58.47	61.73	9.42
M	52.13	54.64	8.40
Median			
Gender	Net Minutes	Gun Minutes	Pace Minutes
F	57.85	61.42	9.32
M	51.38	54.30	8.28
Mode			
Gender	Net Minutes	Gun Minutes	Pace Minutes
F	48.10	61.78	8.97
M	35.47	55.22	7.25
Range			
Gender	Net Minutes	Gun Minutes	Pace Minutes
F	73.85	77.53	11.88
M	71.85	75.32	11.57

# 2006 Pike's Peak 10k Race

## Results: Question 2

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

- While data cleansing, a 'Difference Minutes' variable was created to help analyze the difference between gun time and net time.
- The data showed that the Gun Time for this data was greater than the Net Time. Figure 1 showed an upward trend for racers who have a higher difference in minutes by their placement in the race.
- The Spearman's  $\rho = 0.67$ ,  $p\text{-value} = 0.0000$ , showed a high-positive linear relationship between difference in minutes and race results.

Figure 1: 2006 Pike Peak 10k Race Results  
Difference between Gun/Net Times by Race Place and Gender



\*Difference Minutes = Gun Time - Net Time

# 2006 Pike's Peak 10k Race

## Results: Question 3

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

Top 10 Percentile: Males Aged 40-49 -

Net Minutes =

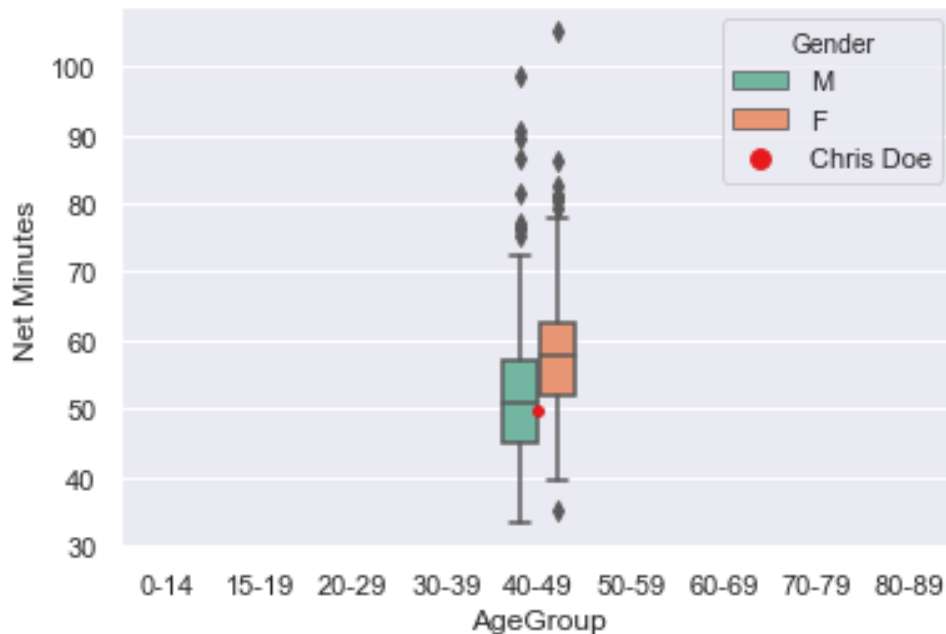
41.69 min

Chris Doe: Net Minutes = 49.72 min

Time difference needed = 8.03 min

- Chris Doe's division consists of racers who are the same gender and part of the same age group.
- This box plot shows the 415 male racers in the 40-49 age group and indicates Chris Doe's Net minutes results in relation to others in his division and compares females of the same age group similarly.
- This boxplot shows that Chris Doe did finish better than the median in his division and the median for females of the same age-group.

Figure 2: Chris Doe: Comparison to Same Division (Ages 40-49)  
(n=698)



Note: # of Racers(40-49 Age Group): Males = 415:Females = 283

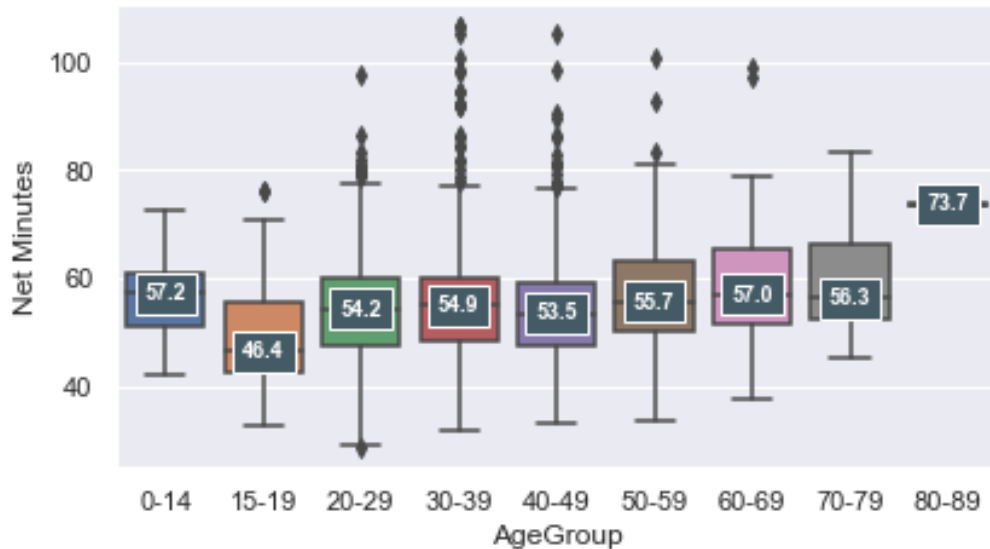
# 2006 Pike's Peak 10k Race

## Results: Question 4

*Compare the race results of each division.*

- This boxplot shown shows the race results in Net Minutes for each age-group.
- The outliers seen on this boxplot is a good indication of the positive skewness we saw in question 1 and supports why the median is a better indication of the central data.
- Racer's aged 15-19 had the lowest median (46.4 min), while racer's aged 80-89 had the highest median (73.7 min).
- The only age-group with an outlier below quartile 1 is the 20-29 age group, with a median of 54.2 min.

Figure 3: 2006 Pike Peak 10k Race Results:  
By Each Division/Age Group



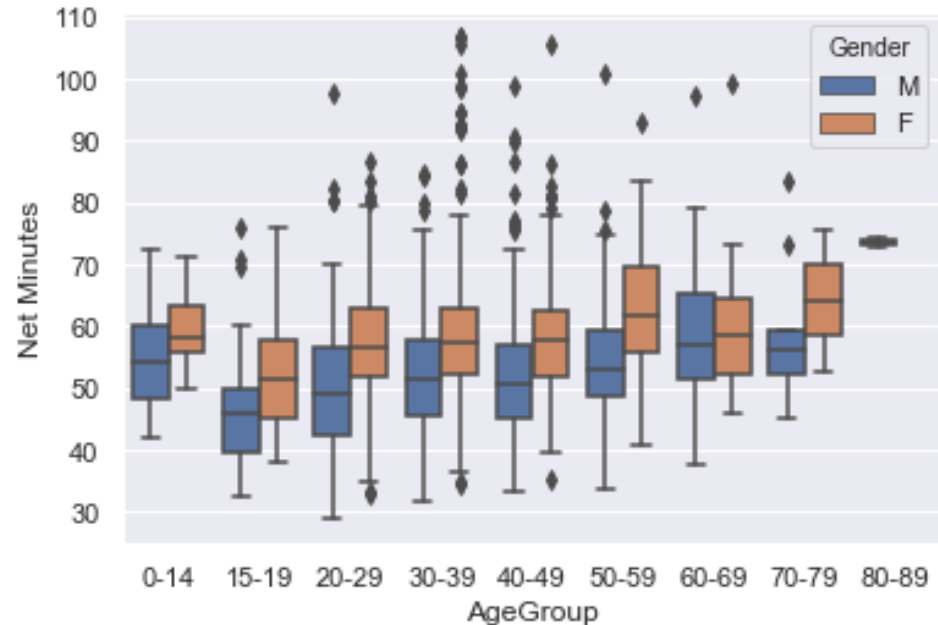
# 2006 Pike's Peak 10k Race

## Results: Question 4 - Part 2

*Compare the race results of each division.*

- After looking at part of each division by age group, we take a closer look at each division with racers of the same age group and gender.
- Both females and males had their best median values in the 15-19 age group.
- By looking at these boxplots we see that the only outliers below quartile 1 are those from age groups 20-29, 30-39, 40-49 and are female, respectively.

Figure 4: 2006 Pike Peak 10k Race Results:  
By Each Division/Age Group and Gender





## 2006 Pike's Peak 10k Race Results: Question 4 - Part 3

*Compare the race results of each division.*

- Figure 5. Shows the Net Time results by Age for all racers.
- Two of the lower peaks come from the the age groups (15-19 and 20-29).
- The standard deviation shown by the shaded areas in Figure 5, indicates that a racer in the 60-69 age group had race results consistent with racers in the top two age groups (15-19, 20-29).



# 2006 Pike's Peak 10k Race

## Results: Question 4 - Part 3

*Compare the race results of each division.*

Figure 5: 2006 Pike Peak 10k Race Results:  
By Divisions/Age Group

