Assignment_8

Zhaowen Fan
&
Rafael Ignacio Gonzalez Chong

UT Austin GSI

Task 1: Calculate the mean, median, first and third quartiles (Q1 and Q3), and standard deviation of height and weight for the individuals in the dataset.

>>> The code:

```
mean_height <- mean(people$height)
mean_weight <- mean(people$weight)

median_height <- median(people$height)
median_weight <- median(people$weight)

quantile_height <- quantile(people$height, probs = c(0.25, 0.75))
quantile_weight <- quantile(people$weight, probs = c(0.25, 0.75))

sd_height <- sd(people$height)
sd_weight <- sd(people$weight)

list(
  Mean_Height = mean_height,
  Median_Height = median_height,
  Q1_Height = quantile_height[1],
  Q3_Height = quantile_height[2],
  SD_Height = sd_height,

  Mean_Weight = mean_weight,
  Median_Weight = median_weight,
  Q1_Weight = quantile_weight[1],
  Q3_Weight = quantile_weight[2],
  SD_Weight = sd_weight
)
```

>>> The output:

```
$Mean_Height
[1] 173.972
$Median_Height
[1] 174
$Q1_Height
25%
164
$Q3_Height
75%
184
$SD_Height
[1] 14.52775
```

```
$Mean_Weight
[1] 89.6396
$Median_Weight
[1] 86.425
$Q1_Weight
  25%
66.38
$Q3_Weight
     75%
109.6675
$SD_Weight
[1] 34.19215
```

Task 2: Generate histograms to visualize the distribution of height and weight for the individuals in the dataset.

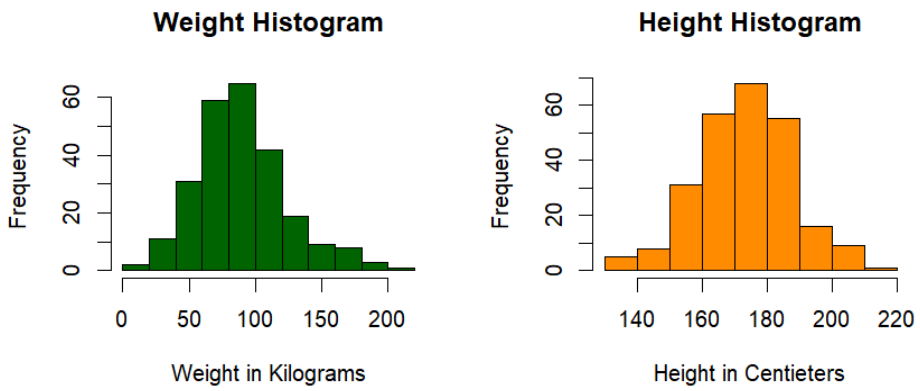
>>> The code:

```
par(mfrow = c(1, 2))

hist(
  people$weight, main = "Weight Histogram",
  xlab = "Weight in Kilograms",
  col = "darkgreen"
)

hist(
  people$height, main = "Height Histogram",
  xlab = "Height in Centieters",
  col = "darkorange"
)
```

>>> The graphs:

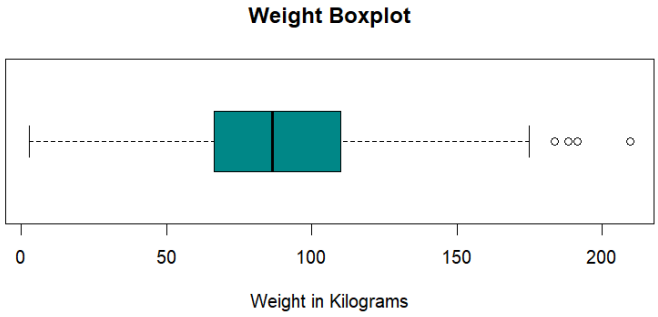


Task 3: Identify any potential outliers in the dataset for weight.
>>> The code:

```
par(mfrow = c(1, 1))

box <- boxplot(
  people$weight, main = "Weight Boxplot",
  col = "darkcyan", horizontal = TRUE,
  xlab = "Weight in Kilograms"
)
cat("There are", length(box$out), "numbers of outliers.")
cat("Outliers are:", box$out)
```

>>> The graphs and outputs:




```
> cat("There are", length(box$out), "numbers of outliers.\n")
There are 4 numbers of outliers.
> cat("Outliers are:", box$out)
Outliers are: 188.3 209.58 191.39 183.59
```

Task 4: Determine the percentage of individuals taller than 150 cm.
>>> The code:

```
tall.dude <- people %>%
  filter(height > 150)

percentage.tall <- nrow(tall.dude) / nrow(people) * 100
paste0(percentage.tall, "%")
```

>>> The output:
The percentage of individuals taller than 150cm is 94.8%.

Task 5: Calculate the percentage of individuals with a weight between 60 kg and 100 kg.
>>> The code:

```
fit.dude <- people %>%
  filter(weight >= 60) %>%
  filter(weight <= 100)

percentage.fit <- nrow(fit.dude) / nrow(people) * 100
paste0(percentage.fit, "%")
```

>>> The output:
The percentage of individuals with a weight between 60kg and 100kg is 49.6%.

Task 6: Identify the weight range that encompasses 60% of the individuals.
>>> The code:

```
quantile.weight <- quantile(people$weight, probs = c(0.2, 0.8))
cat("60% of individuals have weights between", round(quantile.weight[1], 2),
    "kg and", round(quantile.weight[2], 2), "kg.")
```

>>> The output:

```
60% of individuals have weights between 63.7 kg and 114.77 kg.
```

Task 7: Calculate the height above which 1% of the individuals are taller.
>>> The code:

```
super.tall <- quantile(people$height, probs = 0.99)
cat("Only 1% people are taller than", round(super.tall, 2), ".")
```

>>> The output:

```
Only 1% people are taller than 206.02 .
```

Task 8: Determine the weight below which 0.05% of the individuals weigh less.
>>> The code:

```
super.light <- quantile(people$weight, probs = 0.05)
cat("Only 0.05% people are lighter than", round(super.light, 2), ".")
```

>>> The output:

```
Only 0.05% people are lighter than 40.11.
```

Task 9: Compute the BMI for each individual in the dataset and classify those with a BMI greater than 30 as obese.
>>> The code:

```
people$BMI <- people$weight / ((people$height / 100) ^ 2)
summary(people$BMI)

obese <- people %>%
  filter(BMI > 30)
summary(obese$BMI)
print("People obese are")
print(obese)
```

>>> The output:

```
> summary(obese$BMI)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  30.07   32.51   36.32   37.69   41.57   57.06
> print("People obese are")
[1] "People obese are"
> print(obese)
    height weight gender      BMI
1      185 161.64      M 47.22863
2      189 142.74      F 39.95969
3      167 124.38      M 44.59823
4      160 103.50      M 40.42969
5      189 122.31      F 34.24036
6      147  66.34      M 30.70017
7      187 106.90      M 30.56993
8      184 144.36      M 42.63941
9      162  86.47      F 32.94848
10     185 106.78      M 31.19942
11     173  92.87      F 31.03010
12     186 105.21      F 30.41103
13     183 116.91      M 34.90997
14     188 112.78      F 31.90923
15     182 131.50      M 39.69931
16     169 100.28      F 35.11082
17     158  88.47      M 35.43903
18     169 132.15      F 46.26939
19     166 110.17      F 39.98040
20     168 100.05      F 35.44855
21     183 106.83      F 31.90003
22     185 130.98      F 38.27027
23     167 109.83      F 39.38112
24     156  87.24      F 35.84813
25     197 126.28      F 32.53884
26     149  81.68      M 36.79114
27     176 110.58      M 35.69861
```

28	157	110.99	M	45.02820
29	169	89.37	M	31.29092
30	179	165.14	M	51.54021
31	133	67.36	F	38.08016
32	173	117.65	M	39.30970
33	163	91.80	M	34.55155
34	178	140.20	F	44.24946
35	160	85.20	M	33.28125
36	180	119.26	M	36.80864
37	199	188.30	F	47.54930
38	145	93.56	M	44.49941
39	175	96.16	M	31.39918
40	160	118.84	M	46.42187
41	211	172.74	F	38.79967
42	177	104.70	M	33.41952
43	160	79.92	F	31.21875
44	168	90.35	F	32.01176
45	156	75.78	F	31.13905
46	204	174.33	F	41.89014
47	188	143.50	F	40.60095
48	186	150.04	F	43.36918
49	178	138.59	M	43.74132
50	189	110.66	F	30.97898
51	172	100.02	M	33.80882
52	178	116.53	M	36.77882
53	198	209.58	F	53.45883
54	171	88.95	M	30.41962
55	183	159.64	F	47.66938
56	182	101.13	F	30.53073
57	186	107.59	M	31.09897
58	168	109.03	F	38.63024
59	179	99.46	M	31.04148
60	151	95.95	F	42.08149
61	184	104.75	M	30.93986
62	146	72.37	M	33.95102
63	194	134.77	M	35.80880
64	168	130.48	M	46.23016
65	175	135.42	F	44.21878
66	170	97.48	M	33.73010
67	157	99.34	F	40.30184
68	175	174.75	M	57.06122
69	163	108.35	F	40.78061
70	168	94.32	F	33.41837
71	190	166.49	M	46.11911
72	163	92.89	M	34.96180
73	194	136.84	M	36.35881
74	177	112.94	F	36.04967
75	183	112.96	M	33.73048
76	204	125.64	M	30.19031

77	198	121.96	F	31.10907
78	187	145.37	M	41.57111
79	175	147.34	M	48.11102
80	180	122.37	M	37.76852
81	176	110.46	F	35.65987
82	195	143.92	M	37.84878
83	170	106.76	M	36.94118
84	189	191.39	F	53.57913
85	167	83.86	M	30.06920
86	168	110.47	M	39.14045
87	162	96.24	M	36.67124
88	192	162.46	F	44.07010
89	178	131.74	F	41.57935
90	186	109.18	F	31.55856
91	165	98.80	M	36.29017
92	183	124.38	M	37.14055
93	180	132.06	M	40.75926
94	187	183.59	M	52.50079
95	168	90.26	M	31.97988
96	198	122.71	F	31.30038
97	181	111.58	M	34.05879
98	195	115.48	M	30.36949
99	173	116.63	M	38.96889
100	166	95.29	M	34.58049
101	164	118.32	F	43.99167
102	184	107.09	F	31.63103
103	172	100.85	M	34.08937
104	160	89.78	F	35.07031
105	181	119.54	F	36.48851
106	188	114.59	F	32.42134
107	190	160.68	M	44.50970
108	158	81.41	F	32.61096