**Sergio Rosendo**

**COM SCI X 450.2  
Bianca Cung  
Jan 25, 2023**

**Homework 2**

**Task 1:**

**Code:**

bodycat <- c()

bodycat[dataset$bodyfat < 14.00] <- 'Athlete'

bodycat[dataset$bodyfat >= 14.00 & dataset$bodyfat <= 25.00] <- 'Average'

bodycat[dataset$bodyfat > 24] <- 'Obese'

dataset <- cbind(dataset, bodycat)

dataset$bodycat <- factor(dataset$bodycat, levels = c('Athlete', 'Average', 'Obese'))

**Task 2:**

**Code:**summary(dataset$height)

htcat <- c()

htcat[dataset$height < 68.25] <- 'short'

htcat[dataset$height >= 68.25 & dataset$height < 70.00] <- 'below average'

htcat[dataset$height >= 70.00 & dataset$height < 72.25] <- 'above average'

htcat[dataset$height >= 72.25] <- 'tall'

dataset <- cbind(dataset, htcat)

dataset$htcat <- factor(dataset$htcat, levels = c('short', 'below average', 'above average', 'tall'))

**Task 3:  
Code:**

Diagram, schematic

Description automatically generatedbwplot(dataset$weight~dataset$bodycat, xlab = 'Body Category', ylab = 'Body Weight, panel = panel.violin)

**Analysis:**

Based on the shape of the violins in the violin plot we can see that the weights are evenly distributed among the three body types. We can also roughly estimate that ‘Athletes’ are often in the range of 125 to 175 pounds. However, don’t let that mislead you as heigh is also an important factor justifying each weight for these ‘Athletes’.

The violin plot can be misleading and I want to emphasize that height should also be accounted for when determining if someone is overweight.

**Task 4:  
Code:**stem(dataset$weight, scale = 1.0)

11 | 9

12 | 556788

13 | 234466779

14 | 0012345667888999

15 | 001112222233344555556667777888899

16 | 0000011111223333344556667777788888888899

17 | 011111122333334556666777777788888999

18 | 00001122233444444555667788889

19 | 0011122233445566777788999

20 | 0001111223335566778999

21 | 0122366667799999

22 | 335555788

23 | 0345

24 | 1247

25 |

26 | 3

27 |

28 |

29 |

30 |

31 |

32 |

33 |

34 |

35 |

36 | 3

I chose the default scale of 1.0 as I wanted to see the concentration of weights on an interval increasing by 10. The largest concentration goes to people with body weight in the 160s pounds range.  
The top 3 largest concentrations belong to the people with body weight in the range of 140 – 160 pounds range.

**Task 5:**

**Chart, histogram

Description automatically generated**

**Code:**  
hist(dataset$neck[dataset$bodycat == 'Average'], freq = FALSE, main = 'Neck Circumference (cm)' , xlab = 'Neck Circumference (cm)', col = rgb(1.00, 0.00, 0.00, 0.25), xlim = c(31.10, 51.20), ylim = c(0.00, 0.20))

hist(dataset$neck[dataset$bodycat == 'Obese'], freq = FALSE, col = rgb(0.00, 1.00, 0.00, 0.25), add = TRUE)

hist(dataset$neck[dataset$bodycat == 'Athlete'], freq = FALSE, col = rgb(0.00, 0.00, 1.00, 0.25), add = TRUE)

lines(density(dataset$neck[dataset$bodycat == 'Average']), lwd = 1.0)

lines(density(dataset$neck[dataset$bodycat == 'Obese']), lwd = 2.0)

lines(density(dataset$neck[dataset$bodycat == 'Athlete']), lwd = 3.0)

legend('topright', c('Average', 'Obese', 'Athlete'), col = c(rgb(1.00, 0.00, 0.00, 0.25), rgb(0.00, 1.00, 0.00, 0.25), rgb(0.00, 0.00, 1.00, 0.25)), lwd = 10)

**Analysis:**

The histogram above shows us were the largest concentration of neck circumference of all three body type groups. Athletes from about 35.00 – 38.00 cm, average from 36.00 – 38.00 cm and obese 38.00 cm to 42.00 cm

**Task 6:**

**Chart, histogram

Description automatically generatedCode:**

hist(dataset$abdomen[dataset$bodycat == 'Average'], freq = FALSE, main = 'Abdomen Circumference (cm)' , xlab = 'Abdomen Circumference (cm)', col = rgb(1.00, 0.00, 0.00, 0.25), xlim = c(69.40, 148.1), ylim = c(0.00, 0.07))

hist(dataset$abdomen[dataset$bodycat == 'Obese'], freq = FALSE, col = rgb(0.00, 1.00, 0.00, 0.25), add = TRUE)

hist(dataset$abdomen[dataset$bodycat == 'Athlete'], freq = FALSE, col = rgb(0.00, 0.00, 1.00, 0.25), add = TRUE)

lines(density(dataset$abdomen[dataset$bodycat == 'Average']), lwd = 1)

lines(density(dataset$abdomen[dataset$bodycat == 'Obese']), lwd = 2.0)

lines(density(dataset$abdomen[dataset$bodycat == 'Athlete']), lwd = 3.0)

legend('topright', c('Average', 'Obese', 'Athlete'), col = c(rgb(1.00, 0.00, 0.00, 0.25), rgb(0.00, 1.00, 0.00, 0.25), rgb(0.00, 0.00, 1.00, 0.25)), lwd = 10)

**Analysis:**The histogram above shows us were the largest concentration of abdomen circumference of all three body type groups. Athletes from about 70.00 – 90.00 cm, average from 90.00 – 100.00 cm and obese 100.00 cm to 110.00 cm.