Network Code

```
#######################################
# Load and clean the data
#######################################

# Spreadsheet needs to be a .csv file for this code to work
# Be sure to replace "name_of_file" with your actual file name
connect <- read.csv("name_of_file.csv", head=FALSE, stringsAsFactors=FALSE)

# Assign variable names to columns 1 and 2 in the data set
names(connect) <- c("person1","person2")

# Create the connections between people
connect$person1 <- tolower(connect$person1)
connect$person2 <- tolower(connect$person2)
connect$person1 <- gsub(connect$person1, pattern = "-", replacement = " ")
connect$person2 <- gsub(connect$person2, pattern = "-", replacement = " ")

# Find all unique persons in the data set
uni_connect <- c(unique(connect$person1, unique(connect$person2)))

#######################################
# Create the social network!
#######################################

# Create the social network matrix and name it sn_mat
sn_mat <- matrix(0, ncol=length(uni_connect), nrow=length(uni_connect))
for(i in 1:nrow(connect)){
  sn_mat[match(connect[i,1], uni_connect), match(connect[i,2], uni_connect)] = 1
  sn_mat[match(connect[i,2], uni_connect), match(connect[i,1], uni_connect)] = 1
}

# Find and remove the people who have no connections
no_connects <- which(rowSums(sn_mat) == 0)
sn_mat <- sn_mat[-no_connects, -no_connects]
uni_connect <- uni_connect[-no_connects]

#######################################
# Reorder the network by popularity
#######################################

# Order people in the network by how popular they are (i.e. who has the most connections)
# Name this network popular
popular <- data.frame(name = uni_connect, connections = rowSums(sn_mat))
popular <- popular[order(popular$connections, decreasing = TRUE), ]

# Turn the popular network into a social network matrix
pop_mat <- matrix(0, ncol=length(popular$name), nrow=length(popular$name))
for(i in 1:nrow(connect)){
  pop_mat[match(connect[i,1], popular$name), match(connect[i,2], popular$name)] = 1
  pop_mat[match(connect[i,2], popular$name), match(connect[i,1], popular$name)] = 1
}

#######################################
# Make your first network plot! Pretty!
#######################################

# Network plot 1
# Plot should have a dot for each person in the network,
# accompanied by the corresponding person's name
# The lines between different dots represent the connections
set.seed(2)
n1 <- network(pop_mat)
pdf("network_plot1.pdf", width = 36, height = 36)
```

```
plot(n1, vertex.col = c(brewer.pal(9, "Set2")), vertex.border = NA, usearrows = F,
     edge.col = "grey", label = uni_connect, edge.lwd = 0.1)
graphics.off()

########################################
# Determine the typical number
# of connections in the network
########################################

# Histogram of the number of connections each person has
pdf("connections_histogram.pdf")
connect2 <- data.frame(name = uni_connect, connections = rowSums(sn_mat))
ggplot(data = connect2, aes(x = connections)) +
  geom_bar(colour = "white", binwidth = 1) +
  theme_bw() +
  labs(title = "Histogram for number of unique connections",
       x = "Number of Connections", y = "Number of People")
graphics.off()

########################################
# Analyze the most popular people
########################################

# Find the people with the most connections
# NOTE: The value can be increased or decreased.
# If you get less than 5 names returned with the value of "10", decrease the number to 9.
# Continue decreasing until at least 5 names are returned.
# If you get more than 5 names returned with the value of "10", increase the number to 11.
# Continue increasing until 5 or so names are returned.
uni_connect[which(rowSums(sn_mat) > 10)]

# Find the most popular person in the network by determining who has the most connections
uni_connect[which(rowSums(sn_mat) == max(rowSums(sn_mat)))]

# Determine how many connections the most popular person had
max(rowSums(sn_mat))

# Histogram of the number of connections for the most popular people
# NOTE: Use the same value that you found above to return the top 5 names
pdf("popular_histogram.pdf")
connect3 <- subset(connect2, connections >= 10)
ggplot(data = connect3, aes(x = name, y = connections)) +
  theme_bw() +
  geom_bar(stat = "bin") +
  labs(title = "Bargraph of top connected individuals",
       x = "", y = "Number of connections")
graphics.off()

########################################
# Another pretty plot!
########################################

# This plot is the same as the first network plot,
# but identifies the most popular people by a different color
set.seed(2)
n2 <- network(pop_mat)
pdf("network_plot2.pdf", width = 36, height = 36)
plot(n2, vertex.col = c(rep('#E80E84', 5), rep('#76D2E8', 155)),
     vertex.cex = c(rep(2, 5), rep(1, 155)),
     vertex.border=NA, usearrows=F, edge.col="grey", label=popular$name, edge.lwd=0.1)
graphics.off()
```