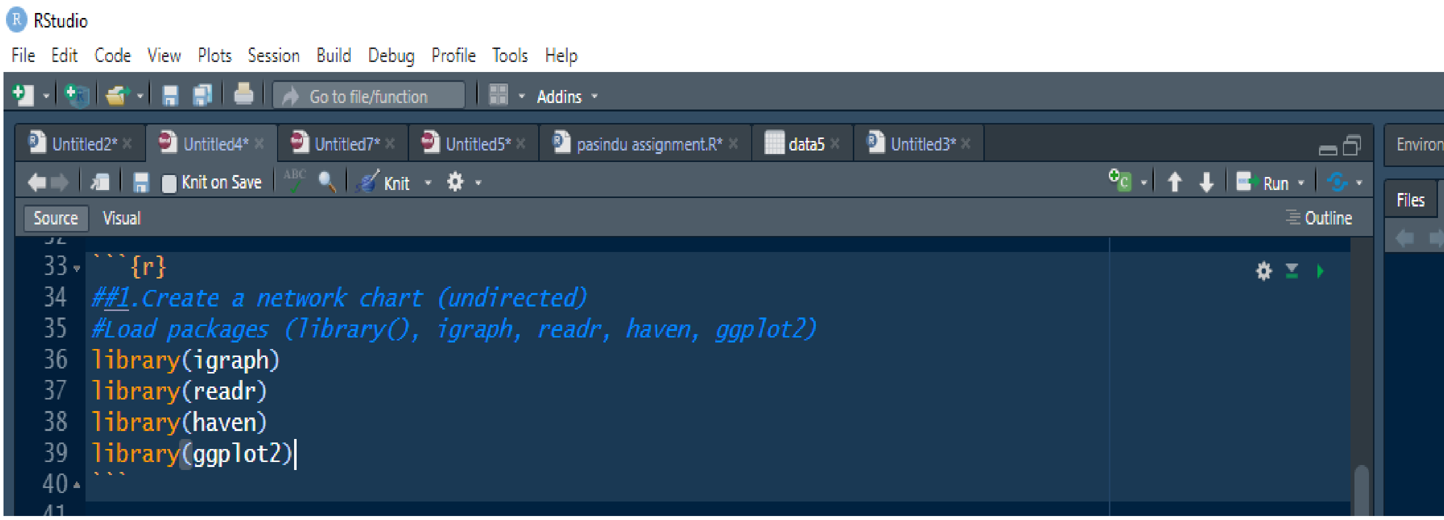
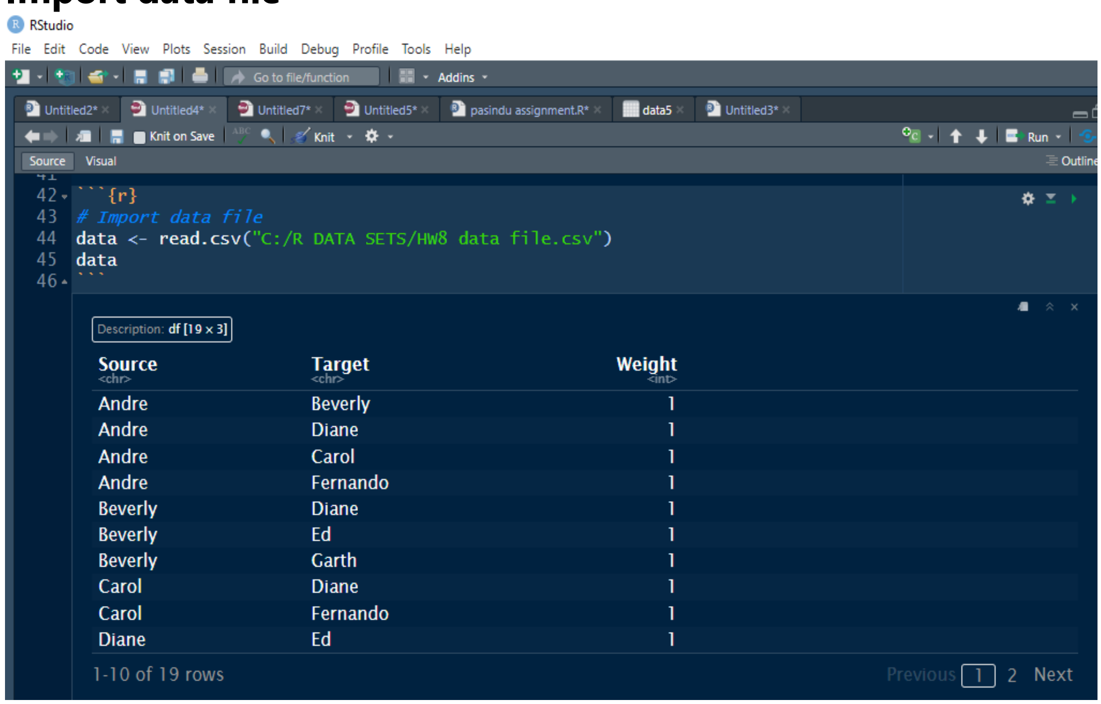
R-studio for Network Analysis

Use my csv file. The data shows the relationships between the two people. Use **Undirect edge** type analysis.

1. Create a network chart (undirected)
    - **Load packages (library(), igraph, readr, haven, ggplot2)**



    - **Import data file**



    -

- Create a graph object (graph.data.frame) – “undirected”

```r
```{r}
# Create a graph object (graph.data.frame) – "undirected"
graph <- graph.data.frame(data, directed = FALSE)
graph
```

IGRAPH bc27331 UN-- 11 19 --
+ attr: name (v/c), weight (e/n)
+ edges from bc27331 (vertex names):
 [1] Andre   --Beverly Andre   --Diane   Andre   --Carol   Andre   --Fernando
 [5] Beverly --Diane   Beverly --Ed   Beverly --Garth   Carol   --Diane
 [9] Carol   --Fernando Diane   --Ed   Diane   --Fernando Diane   --Garth
[13] Ed   --Garth   Fernando--Garth   Fernando--Heather   Garth   --Heather
[17] Heather --Ike   Ike   --Jane   Heather --Kathy
```

2. Modify the layout of the network chart.

set.seed(1234)

```r
```{r}
## 2.Modify the layout of the network chart.
# Set seed
set.seed(1234)
```
```

- fruchterman.reingold layout

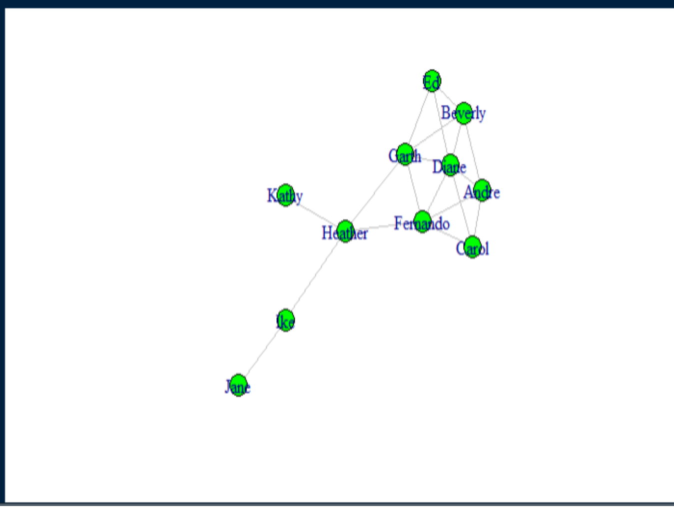
```r
```{r}
# Define layout ofruchterman.reingold layout
layout <- layout.fruchterman.reingold(graph)

```
```

- vertex color – “green”
- edge color – “grey”

Take a Screenshot of the outcome

```r
# Plot network with modified layout and colors
plot(g, layout = layout, vertex.color = "green", edge.color = "grey")
```



3. Calculate and Answer the following the following information
   - Calcuate **average path length** (average geodesic distance). What does this measure mean?

```r
##3.Calculate and Answer the following the following information
# o Calcuate average path length (average geodesic distance). What does this measure mean?
avg_path_length <- mean_distance(graph)
cat("Average path length:", avg_path_length, "\n")
# The average path length is a measure of the average distance between any two nodes in a network. It reflects the efficiency of communication in the network and can be used to evaluate the network's ability to transfer information or resources. A lower average path length indicates a more connected and efficient network, while a higher average path length indicates a less efficient network with less communication between nodes.
```

```
Average path length: 2.072727
```

   - Calculate the **degree centrality**

```r
# Calculate degree centrality
degree_centrality <- degree(graph)
cat("Degree centrality:", degree_centrality, "\n")
```

```
Degree centrality: 4 4 3 6 3 5 5 4 2 1 1
```

- Create a data frame using the degree centrality (*Hint: as.data.frame)

```{r}
# Create data frame using degree centrality
degree_df <- as.data.frame(degree_centrality)
degree_df
```

Description: df [11 x 1]

	degree_centrality <dbl>
Andre	4
Beverly	4
Carol	3
Diane	6
Ed	3
Fernando	5
Garth	5
Heather	4
Ike	2
Jane	1

1-10 of 11 rows Previous 1 2 Next

- Identify the **minimum** degree, **maximum** degree, and **average degree**

```{r}
# Identify minimum, maximum, and average degree
min_degree <- min(degree_centrality)
max_degree <- max(degree_centrality)
avg_degree <- mean(degree_centrality)
cat("Minimum degree:", min_degree, "\n")
cat("Maximum degree:", max_degree, "\n")
cat("Average degree:", avg_degree, "\n")

# What is the difference between average path length and the average degree?
#Why is the result different?
#The average path length measures the average number of steps along the shortest paths for all possible pairs of nodes in a network, providing a measure of the network's efficiency of communication.On the other hand, the average degree measures the average number of connections that a node has in a network, providing a measure of the network's density or connectedness. These two measures capture different aspects of the network structure, and their values may not be strongly correlated. The difference in their values is due to the fact that they are measuring different aspects of the network.


```

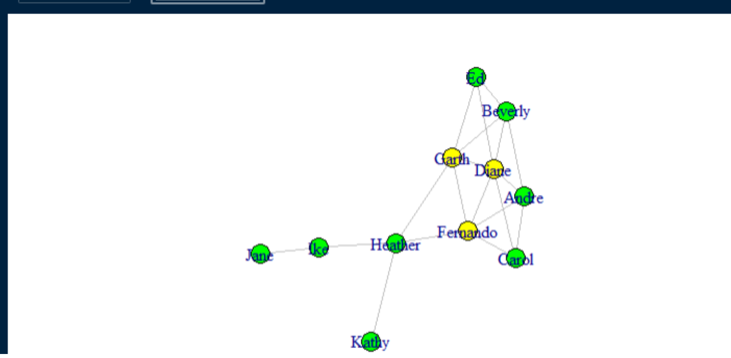
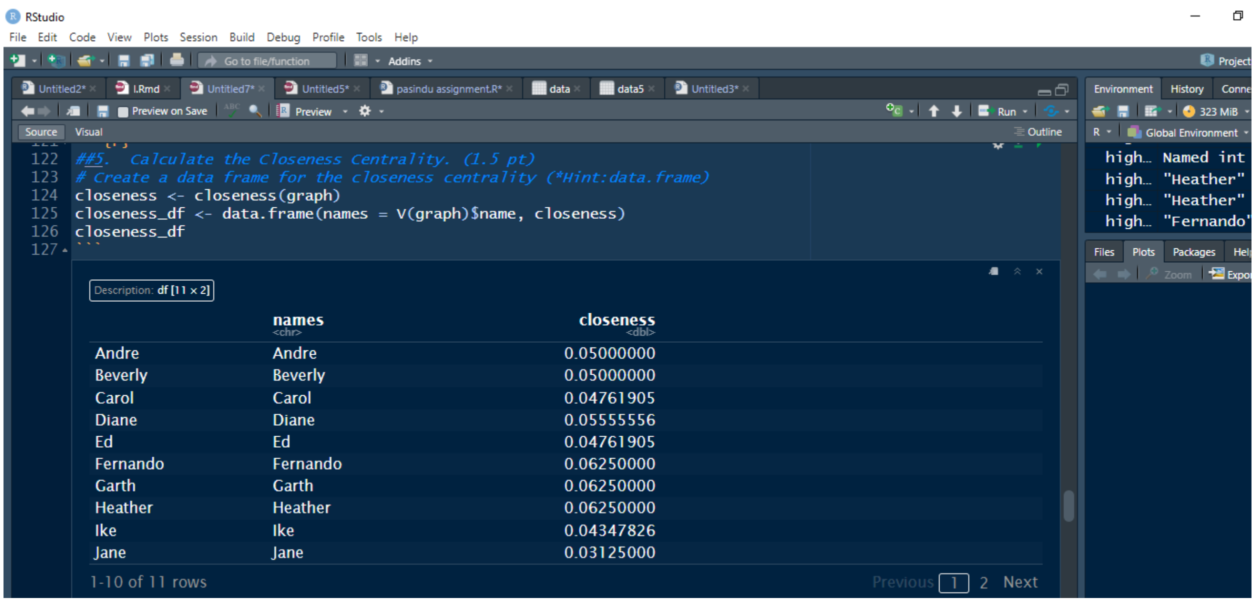
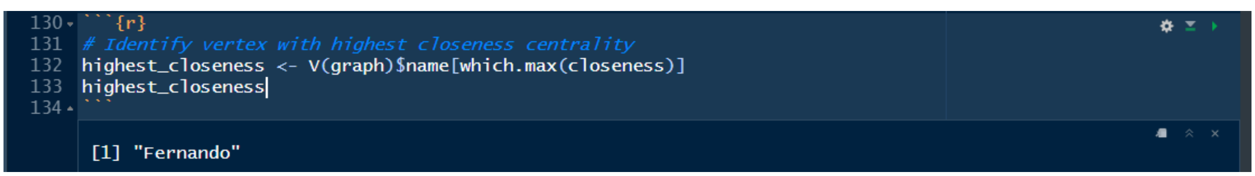
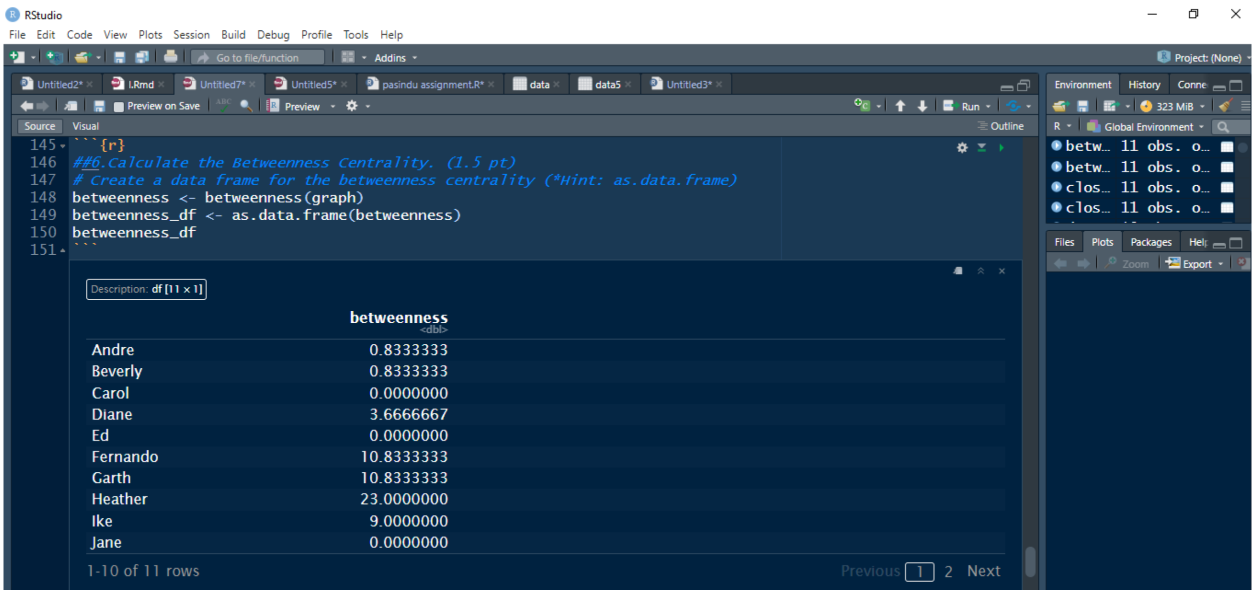
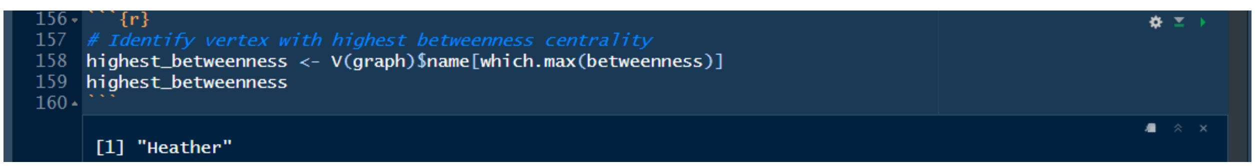
```
Minimum degree: 1
Maximum degree: 6
Average degree: 3.454545
```

- What is the difference between **average path length** and the **average degree**? Why is the result different?

#The average path length measures the average number of steps along the shortest paths for all possible pairs of nodes in a network, providing a measure of the network's efficiency of communication. On the other hand, the average degree measures the average number of connections that a node has in a network, providing a measure of the network's density or connectedness. These two measures capture different aspects of the network structure, and their values may not be strongly correlated. The difference in their values is due to the fact that they are measuring different aspects of the network.

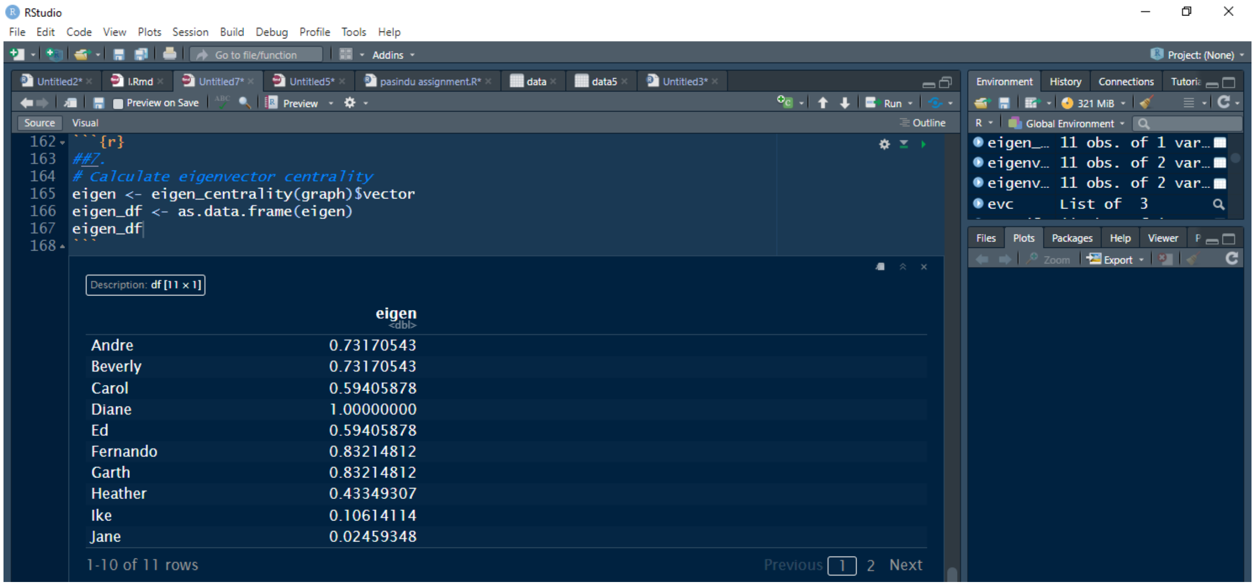
4. Modify the layout of the network chart by degree centrality.

```
```{r}
##4.Modify the layout of the network chart by degree centrality. Specifically, change the color of the vertex
with degree centrality of higher than 4 to Yellow.
# Calculate the degree centrality
degree_centrality <- degree(graph)
degree_centrality
```
```

```
  Andre  Beverly     Carol     Diane        Ed Fernando     Garth  Heather       Ike      Jane     Kathy
      4        4         3         6         3        5         5        4         2         1         1
```

Specifically, change the color of the vertex with degree centrality of <ins>**higher than 4**</ins> to **Yellow.**

```
node_colors <- ifelse(degree_centrality > 4, "yellow", "green")
node_colors
# Modify the layout of the network chart
plot(graph, layout = layout.fruchterman.reingold, vertex.color = node_colors, edge.color = "grey")
```
```

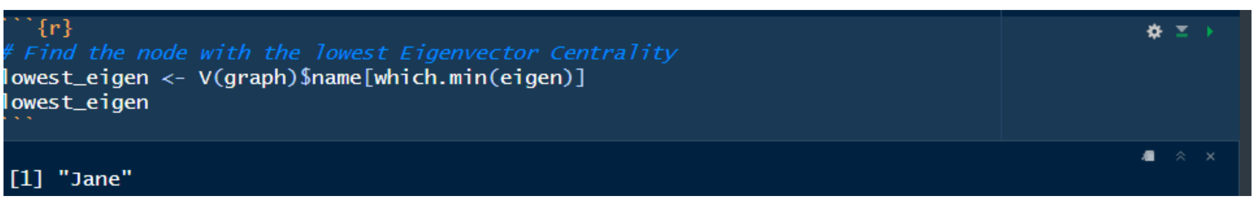

```

5. Calculate the **Closeness Centrality**. **(1.5 pt)**
    - Create a data frame for the closeness centrality (**Hint:*data.frame)



    - Who has the highest Closeness Centrality?



6. Calculate the **Betweenness Centrality**. **(1.5 pt)**
    - Create a data frame for the betweenness centrality (**Hint*: as.data.frame)



````
```{r}
##6.Calculate the Betweenness Centrality. (1.5 pt)
# Create a data frame for the betweenness centrality (*Hint: as.data.frame)
betweenness <- betweenness(graph)
betweenness_df <- as.data.frame(betweenness)
betweenness_df
```
````

Description: df [11 x 1]

	betweenness <dbl>
Andre	0.8333333
Beverly	0.8333333
Carol	0.0000000
Diane	3.6666667
Ed	0.0000000
Fernando	10.8333333
Garth	10.8333333
Heather	23.0000000
Ike	9.0000000
Jane	0.0000000

1-10 of 11 rows Previous 1 2 Next

    - Who has the highest Betweenness Centrality?



````
```{r}
# Identify vertex with highest betweenness centrality
highest_betweenness <- V(graph)$name[which.max(betweenness)]
highest_betweenness
```
````

```
[1] "Heather"
```

7. Calculate the **Eigenvector Centrality. (1.5 pt)**
    - Create a data frame for the betweenness centrality (**Hint*: as.data.frame)



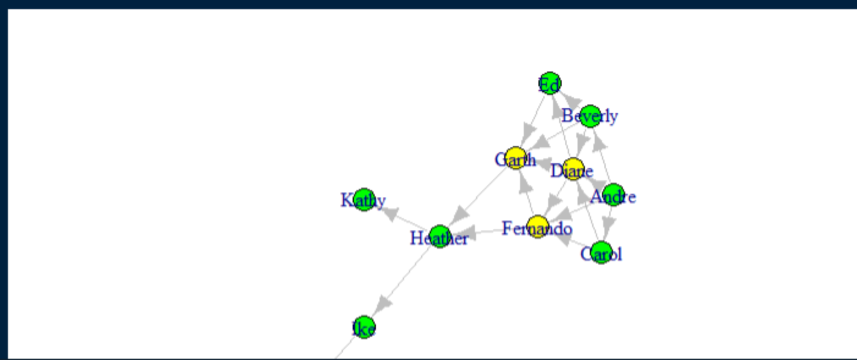
- Who has the lowest Eigenvector Centrality?



8. Repeat the Q1-Q11 process. But this time, create a <u>**DIRECTED**</u> network chart (**Hint*: Do not need to re-write the entire thing. Only the assignment, in the beginning, needs to change!)
    - How is the result different from the <u>**Undirected**</u> network? Choose one of the centrality metrics to look into (e.g., degree centrality, closeness centrality, betweenness centrality, eigenvector centrality). Is the result different from the Undirected network?

````r
graph <- graph.data.frame(data, directed = TRUE)
graph
set.seed(1234)
plot(graph, layout = layout_with_fr(graph),
     vertex.color = ifelse(degree(graph) > 4, "yellow", "green"),
     edge.color = "grey")
```
````

R Console




```{r}
dc <- degree(graph, mode = "in")
dc_df <- data.frame(name = V(graph)$name, degree_centrality = dc)
dc_df
```

Description: df [11 x 2]

	name <chr>	degree_centrality <dbl>
Andre	Andre	0
Beverly	Beverly	1
Carol	Carol	1
Diane	Diane	3
Ed	Ed	2
Fernando	Fernando	3
Garth	Garth	4
Heather	Heather	2
Ike	Ike	1
Jane	Jane	1

1-10 of 11 rows Previous 1 2 Next

#If we compare this to the degree centrality in the undirected network, we can see that the values are different. In the directed network, we have separate measures for the in-degree and out-degree of each vertex, which can result in different values than in the undirected case.

- Based on the results, would your decision of the influencer change from Q12?
  Take a Screenshot of the outcome

#Based on the results, our decision on the influencer might change depending on the centrality metric we are using. For example, if we focus on in-degree centrality, we can see that vertex 1 has the highest value, which indicates that it receives the most incoming
```

connections. However, if we look at betweenness centrality, vertex 5 has the highest value, which indicates that it plays a critical role in connecting different parts of the network. #Therefore, it is important to carefully choose the appropriate centrality metric depending on the research question and the characteristics of the network.