Final Project

SDS348 Spring 2021

Name: Santhosh Saravanan

EID: sks3648

```
library(readxl)
library(tidyverse)
library(kableExtra)
USArrests6_1973 <- as.data.frame(read_csv("~/git/SDS348/Project/Datasets/USArrests6_1973.csv"))
glimpse(USArrests6_1973)
```

```
## Rows: 50
## Columns: 5
## $ X1       <chr> "Alabama", "Alaska", "Arizona", "Arkansas", "California", "Co…
## $ Murder   <dbl> 13.2, 10.0, 8.1, 8.8, 9.0, 7.9, 3.3, 5.9, 15.4, 17.4, 5.3, 2.…
## $ Assault  <dbl> 236, 263, 294, 190, 276, 204, 110, 238, 335, 211, 46, 120, 24…
## $ UrbanPop <dbl> 58, 48, 80, 50, 91, 78, 77, 72, 80, 60, 83, 54, 83, 65, 57, 6…
## $ Rape     <dbl> 21.2, 44.5, 31.0, 19.5, 40.6, 38.7, 11.1, 15.8, 31.9, 25.8, 2…
```

```
allArrests <- as.data.frame(read_csv("~/git/SDS348/Project/Datasets/state_crime_all.csv"))
glimpse(allArrests)
```

```
## Rows: 2,751
## Columns: 21
## $ State                      <chr> "Alabama", "Alabama", "Alabama", "Alabam…
## $ Year                       <dbl> 1960, 1961, 1962, 1963, 1964, 1965, 1966…
## $ Data.Population            <dbl> 3266740, 3302000, 3358000, 3347000, 3407…
## $ Data.Rates.Property.All    <dbl> 1035.4, 985.5, 1067.0, 1150.9, 1358.7, 1…
## $ Data.Rates.Property.Burglary <dbl> 355.9, 339.3, 349.1, 376.9, 466.6, 473.7…
## $ Data.Rates.Property.Larceny  <dbl> 592.1, 569.4, 634.5, 683.4, 784.1, 812.1…
## $ Data.Rates.Property.Motor    <dbl> 87.3, 76.8, 83.4, 90.6, 108.0, 106.9, 13…
## $ Data.Rates.Violent.All     <dbl> 186.6, 168.5, 157.3, 182.7, 213.1, 199.8…
## $ Data.Rates.Violent.Assault <dbl> 138.1, 128.9, 119.0, 142.1, 163.0, 149.1…
## $ Data.Rates.Violent.Murder  <dbl> 12.4, 12.9, 9.4, 10.2, 9.3, 11.4, 10.9, …
## $ Data.Rates.Violent.Rape    <dbl> 8.6, 7.6, 6.5, 5.7, 11.7, 10.6, 9.7, 10.…
## $ Data.Rates.Violent.Robbery <dbl> 27.5, 19.1, 22.5, 24.7, 29.1, 28.7, 32.0…
## $ Data.Totals.Property.All   <dbl> 33823, 32541, 35829, 38521, 46290, 48215…
## $ Data.Totals.Property.Burglary <dbl> 11626, 11205, 11722, 12614, 15898, 16398…
## $ Data.Totals.Property.Larceny  <dbl> 19344, 18801, 21306, 22874, 26713, 28115…
## $ Data.Totals.Property.Motor    <dbl> 2853, 2535, 2801, 3033, 3679, 3702, 4606…
## $ Data.Totals.Violent.All    <dbl> 6097, 5564, 5283, 6115, 7260, 6916, 8098…
## $ Data.Totals.Violent.Assault <dbl> 4512, 4255, 3995, 4755, 5555, 5162, 6249…
## $ Data.Totals.Violent.Murder <dbl> 406, 427, 316, 340, 316, 395, 384, 415, …
## $ Data.Totals.Violent.Rape   <dbl> 281, 252, 218, 192, 397, 367, 341, 371, …
## $ Data.Totals.Violent.Robbery <dbl> 898, 630, 754, 828, 992, 992, 1124, 1167…
```

```
colnames(USArrests6_1973)[1] <- "State"
allArrests1973 <- allArrests %>% filter(Year == 1973 & State != "United States")
USArrestscombined <- USArrests6_1973 %>% left_join(allArrests1973,by = c("State")) %>% filter(!is.na(State))
NE<- c("Connecticut","Maine","Massachusetts","New Hampshire",
            "Rhode Island","Vermont","New Jersey","New York",
            "Pennsylvania")
MW<- c("Indiana","Illinois","Michigan","Ohio","Wisconsin",
            "Iowa","Kansas","Minnesota","Missouri","Nebraska",
            "North Dakota","South Dakota")
S<- c("Delaware","District of Columbia","Florida","Georgia",
          "Maryland","North Carolina","South Carolina","Virginia",
          "West Virginia","Alabama","Kentucky","Mississippi",
          "Tennessee","Arkansas","Louisiana","Oklahoma","Texas")
W<- c("Arizona","Colorado","Idaho","New Mexico","Montana",
          "Utah","Nevada","Wyoming","Alaska","California",
          "Hawaii","Oregon","Washington")
USArrestscombined <- USArrestscombined %>%
  mutate(Region = case_when(State %in% MW ~ "MidWest",
                            State %in% W  ~ "West",
                            State %in% NE ~ "NorthEast",
                            State %in% S ~ "South")) %>% arrange(desc(Region))
USArrestscombinedImportant <- USArrestscombined %>% select(-c(Data.Rates.Violent.Murder, Data.Rates.Violent.Assault, Data.Rates.Violent.Rape))
```

```
USArrestscombinedImportant %>%
  summarise(mean_rape = mean(Rape, na.rm = TRUE),mean_assault = mean(Assault, na.rm = TRUE),mean_murder = mean(Mu
rder, na.rm = TRUE), mean_burglary = mean(Data.Rates.Property.Burglary,na.rm = TRUE), mean_larceny = mean(Data.Ra
tes.Property.Larceny,na.rm = TRUE), mean_motor = mean(Data.Rates.Property.Motor, na.rm = TRUE)) %>% kbl(caption =
"Recreating booktabs style table") %>%
  kable_classic(full_width = F, html_font = "Cambria") %>% kable_material_dark()
```

Recreating booktabs style table

mean_rape	mean_assault	mean_murder	mean_burglary	mean_larceny	mean_motor
21.232	170.76	7.788	1096.822	2041.032	369.798

```
USArrestscombinedImportant %>%
  summarise(min_rape = min(Rape, na.rm = TRUE),min_assault = min(Assault, na.rm = TRUE),min_murder = min(Murder,
na.rm = TRUE), min_burglary = min(Data.Rates.Property.Burglary,na.rm = TRUE), min_larceny = min(Data.Rates.Proper
ty.Larceny,na.rm = TRUE), min_motor = min(Data.Rates.Property.Motor, na.rm = TRUE)) %>% kbl(caption = "Recreating
booktabs style table") %>%
  kable_classic(full_width = F, html_font = "Cambria") %>% kable_material_dark()
```

Recreating booktabs style table

min_rape	min_assault	min_murder	min_burglary	min_larceny	min_motor
7.3	45	0.8	383.4	824.9	107.1

```
USArrestscombinedImportant %>%
  summarise(max_rape = max(Rape, na.rm = TRUE),max_assault = max(Assault, na.rm = TRUE),max_murder = max(Murder,
na.rm = TRUE), max_burglary = max(Data.Rates.Property.Burglary,na.rm = TRUE), max_larceny = max(Data.Rates.Proper
ty.Larceny,na.rm = TRUE), max_motor = max(Data.Rates.Property.Motor, na.rm = TRUE)) %>% kbl(caption = "Recreating
booktabs style table") %>%
  kable_classic(full_width = F, html_font = "Cambria") %>% kable_material_dark()
```

Recreating booktabs style table

max_rape	max_assault	max_murder	max_burglary	max_larceny	max_motor
46	337	17.4	2149.8	3720.1	1109.6

```
USArrestscombinedImportant %>%
  summarise(sd_rape = sd(Rape, na.rm = TRUE),sd_assault = sd(Assault, na.rm = TRUE),sd_murder = sd(Murder, na.rm
= TRUE), sd_burglary = sd(Data.Rates.Property.Burglary,na.rm = TRUE), sd_larceny = sd(Data.Rates.Property.Larcen
y,na.rm = TRUE), sd_motor = sd(Data.Rates.Property.Motor, na.rm = TRUE)) %>% kbl(caption = "Recreating booktabs s
tyle table") %>%
  kable_classic(full_width = F, html_font = "Cambria") %>% kable_material_dark()
```

Recreating booktabs style table

sd_rape	sd_assault	sd_murder	sd_burglary	sd_larceny	sd_motor
9.366385	83.33766	4.35551	405.4954	671.5986	200.4927

```
USArrestscombinedImportant %>%
  summarise(variation_rape = var(Rape, na.rm = TRUE),variation_assault = var(Assault, na.rm = TRUE),variation_mur
der = var(Murder, na.rm = TRUE), variation_burglary = var(Data.Rates.Property.Burglary,na.rm = TRUE), variation_l
arceny = var(Data.Rates.Property.Larceny,na.rm = TRUE), variation_motor = var(Data.Rates.Property.Motor, na.rm =
TRUE)) %>% kbl(caption = "Recreating booktabs style table") %>%
  kable_classic(full_width = F, html_font = "Cambria") %>% kable_material_dark()
```

Recreating booktabs style table

variation_rape	variation_assault	variation_murder	variation_burglary	variation_larceny	variation_motor
87.72916	6945.166	18.97047	164426.5	451044.7	40197.32

```
USArrestscombinedImportant %>%
  summarise(Distinct_Values_rape = n_distinct(Rape, na.rm = TRUE),Distinct_Values_assault = n_distinct(Assault, n
a.rm = TRUE),Distinct_Values_murder = n_distinct(Murder, na.rm = TRUE), Distinct_Values_burglary = n_distinct(Dat
a.Rates.Property.Burglary,na.rm = TRUE), Distinct_Values_larceny = n_distinct(Data.Rates.Property.Larceny,na.rm =
TRUE), Distinct_Values_motor = n_distinct(Data.Rates.Property.Motor, na.rm = TRUE)) %>% kbl(caption = "Recreating
booktabs style table") %>%
  kable_classic(full_width = F, html_font = "Cambria") %>% kable_material_dark()
```

Recreating booktabs style table

Distinct_Values_rape	Distinct_Values_assault	Distinct_Values_murder	Distinct_Values_burglary	Distinct_Values_larceny	Distinct_Values_motor
48	45	43	50	50	50

```
USArrestscombinedImportant %>%
  summarise(Quantile_rape = quantile(Rape, na.rm = TRUE),Quantile_assault = quantile(Assault, na.rm = TRUE),Quant
ile_murder = quantile(Murder, na.rm = TRUE), Quantile_burglary = quantile(Data.Rates.Property.Burglary,na.rm = TR
UE), Quantile_larceny = quantile(Data.Rates.Property.Larceny,na.rm = TRUE), Quantile_motor = quantile(Data.Rates.
Property.Motor, na.rm = TRUE)) %>% kbl(caption = "Recreating booktabs style table") %>%
  kable_classic(full_width = F, html_font = "Cambria") %>% kable_material_dark() %>% footnote(general = "Here are
general comments of the table. ",
           number = c("First value for column  is at 0%; ", "Second value of for column is at 25%"),
           alphabet = c("Third value for column is at 50%; ", "Fourth value for column  is at 75%; "),
           symbol = c("Fifth value for column  is at 100% ")
           )
```

Recreating booktabs style table

Quantile_rape	Quantile_assault	Quantile_murder	Quantile_burglary	Quantile_larceny	Quantile_motor
7.300	45	0.800	383.400	824.900	107.100
15.075	109	4.075	813.325	1494.775	217.350
20.100	159	7.250	1027.200	2004.500	340.800
26.175	249	11.250	1274.050	2509.275	502.475
46.000	337	17.400	2149.800	3720.100	1109.600

Note:

Here are general comments of the table.

[1] First value for column is at 0%;

[2] Second value of for column is at 25%

[a] Third value for column is at 50%;

[b] Fourth value for column is at 75%;

[*] Fifth value for column is at 100%

```
USArrestscombinedImportant %>%
  group_by(Region) %>%
  summarise(mean_rape = mean(Rape, na.rm = TRUE),mean_assault = mean(Assault, na.rm = TRUE),mean_murder = mean(Mu
rder, na.rm = TRUE), mean_burglary = mean(Data.Rates.Property.Burglary,na.rm = TRUE), mean_larceny = mean(Data.Ra
tes.Property.Larceny,na.rm = TRUE), mean_motor = mean(Data.Rates.Property.Motor, na.rm = TRUE)) %>% kbl(caption =
"Recreating booktabs style table") %>%
  kable_classic(full_width = F, html_font = "Cambria") %>% kable_material_dark()
```

Recreating booktabs style table

Region	mean_rape	mean_assault	mean_murder	mean_burglary	mean_larceny	mean_motor

Region	mean_rape	mean_assault	mean_murder	mean_burglary	mean_larceny	mean_motor
MidWest	18.44167	120.3333	5.700000	894.525	1972.400	317.7750
NorthEast	13.77778	126.6667	4.700000	1041.656	1677.667	508.7333
South	21.16250	220.0000	11.706250	1025.006	1657.412	293.0000
West	29.05385	187.2308	7.030769	1410.138	2828.092	416.1538

```
USArrestscombinedImportant %>%
  group_by(Region) %>%
  summarise(min_rape = min(Rape, na.rm = TRUE),min_assault = min(Assault, na.rm = TRUE),min_murder = min(Murder,
na.rm = TRUE), min_burglary = min(Data.Rates.Property.Burglary,na.rm = TRUE), min_larceny = min(Data.Rates.Proper
ty.Larceny,na.rm = TRUE), min_motor = min(Data.Rates.Property.Motor, na.rm = TRUE)) %>% kbl(caption = "Recreating
booktabs style table") %>%
  kable_classic(full_width = F, html_font = "Cambria") %>% kable_material_dark()
```

Recreating booktabs style table

Region	min_rape	min_assault	min_murder	min_burglary	min_larceny	min_motor
MidWest	7.3	45	0.8	383.4	1406.0	131.4
NorthEast	7.8	48	2.1	685.0	1090.7	136.2
South	9.3	81	5.7	415.8	824.9	107.1
West	14.2	46	2.6	699.7	2237.8	207.0

```
USArrestscombinedImportant %>%
  group_by(Region) %>%
  summarise(max_rape = max(Rape, na.rm = TRUE),max_assault = max(Assault, na.rm = TRUE),max_murder = max(Murder,
na.rm = TRUE), max_burglary = max(Data.Rates.Property.Burglary,na.rm = TRUE), max_larceny = max(Data.Rates.Proper
ty.Larceny,na.rm = TRUE), max_motor = max(Data.Rates.Property.Motor, na.rm = TRUE)) %>% kbl(caption = "Recreating
booktabs style table") %>%
  kable_classic(full_width = F, html_font = "Cambria") %>% kable_material_dark()
```

Recreating booktabs style table

Region	max_rape	max_assault	max_murder	max_burglary	max_larceny	max_motor
MidWest	35.1	255	12.1	1584.6	2771.3	548.3
NorthEast	26.1	254	11.1	1348.2	2312.3	1109.6
South	31.9	337	17.4	1857.2	3048.6	547.8
West	46.0	294	12.2	2149.8	3720.1	635.9

```
USArrestscombinedImportant %>%
  group_by(Region) %>%
  summarise(sd_rape = sd(Rape, na.rm = TRUE),sd_assault = sd(Assault, na.rm = TRUE),sd_murder = sd(Murder, na.rm
= TRUE), sd_burglary = sd(Data.Rates.Property.Burglary,na.rm = TRUE), sd_larceny = sd(Data.Rates.Property.Larcen
y,na.rm = TRUE), sd_motor = sd(Data.Rates.Property.Motor, na.rm = TRUE)) %>% kbl(caption = "Recreating booktabs s
tyle table") %>%
  kable_classic(full_width = F, html_font = "Cambria") %>% kable_material_dark()
```

Recreating booktabs style table

Region	sd_rape	sd_assault	sd_murder	sd_burglary	sd_larceny	sd_motor
MidWest	7.981736	71.53935	3.558345	337.1583	354.8975	139.8711
NorthEast	5.942806	64.85754	3.047950	250.4281	351.4370	334.4643

Region	sd_rape	sd_assault	sd_murder	sd_burglary	sd_larceny	sd_motor
South	5.627536	74.20782	3.760934	332.2478	626.5774	129.6427
West	10.997774	80.32761	3.062511	480.6029	426.8271	152.4576

```
USArrestscombinedImportant %>%
  group_by(Region) %>%
  summarise(variation_rape = var(Rape, na.rm = TRUE),variation_assault = var(Assault, na.rm = TRUE),variation_mur
der = var(Murder, na.rm = TRUE), variation_burglary = var(Data.Rates.Property.Burglary,na.rm = TRUE), variation_l
arceny = var(Data.Rates.Property.Larceny,na.rm = TRUE), variation_motor = var(Data.Rates.Property.Motor, na.rm =
TRUE)) %>% kbl(caption = "Recreating booktabs style table") %>%
  kable_classic(full_width = F, html_font = "Cambria") %>% kable_material_dark()
```

Recreating booktabs style table

Region	variation_rape	variation_assault	variation_murder	variation_burglary	variation_larceny	variation_motor
MidWest	63.70811	5117.879	12.661818	113675.70	125952.3	19563.93
NorthEast	35.31694	4206.500	9.290000	62714.25	123508.0	111866.37
South	31.66917	5506.800	14.144625	110388.59	392599.3	16807.23
West	120.95103	6452.526	9.378974	230979.14	182181.4	23243.33

```
USArrestscombinedImportant %>%
  group_by(Region) %>%
  summarise(Distinct_Values_rape = n_distinct(Rape, na.rm = TRUE),Distinct_Values_assault = n_distinct(Assault, n
a.rm = TRUE),Distinct_Values_murder = n_distinct(Murder, na.rm = TRUE), Distinct_Values_burglary = n_distinct(Dat
a.Rates.Property.Burglary,na.rm = TRUE), Distinct_Values_larceny = n_distinct(Data.Rates.Property.Larceny,na.rm =
TRUE), Distinct_Values_motor = n_distinct(Data.Rates.Property.Motor, na.rm = TRUE)) %>% kbl(caption = "Recreating
booktabs style table") %>%
  kable_classic(full_width = F, html_font = "Cambria") %>% kable_material_dark()
```

Recreating booktabs style table

Region	Distinct_Values_rape	Distinct_Values_assault	Distinct_Values_murder	Distinct_Values_burglary	Distinct_Values_larceny	Distinct
MidWest	12	12	12	12	12	
NorthEast	9	9	8	9	9	
South	16	16	14	16	16	
West	13	12	13	13	13	

```
USArrestscombinedImportant %>%
  group_by(Region) %>%
  summarise(Quantile_rape = quantile(Rape, na.rm = TRUE),Quantile_assault = quantile(Assault, na.rm = TRUE),Quant
ile_murder = quantile(Murder, na.rm = TRUE), Quantile_burglary = quantile(Data.Rates.Property.Burglary,na.rm = TR
UE), Quantile_larceny = quantile(Data.Rates.Property.Larceny,na.rm = TRUE), Quantile_motor = quantile(Data.Rates.
Property.Motor, na.rm = TRUE)) %>% kbl(caption = "Recreating booktabs style table") %>%
  kable_classic(full_width = F, html_font = "Cambria") %>% kable_material_dark() %>% footnote(general = "Here are
general comments of the table. ",
          number = c("First value of region  for column  is at 0%; ", "Second value of region  for column is at
25%"),
          alphabet = c("Third value of region for column is at 50%; ", "Fourth value of region  for column  is a
t 75%; "),
          symbol = c("Fifth value of region  for column  is at 100% ")
          )
```

Recreating booktabs style table

Region	Quantile_rape	Quantile_assault	Quantile_murder	Quantile_burglary	Quantile_larceny	Quantile_motor
MidWest	7.300	45.0	0.800	383.400	1406.000	131.400

Region	Quantile_rape	Quantile_assault	Quantile_murder	Quantile_burglary	Quantile_larceny	Quantile_motor
MidWest	12.425	68.0	2.675	636.550	1834.650	205.550
MidWest	17.250	107.5	5.150	979.700	2024.000	319.700
MidWest	22.050	134.5	7.725	1031.825	2071.975	403.775
MidWest	35.100	255.0	12.100	1584.600	2771.300	548.300
NorthEast	7.800	48.0	2.100	685.000	1090.700	136.200
NorthEast	9.500	83.0	2.200	801.500	1489.900	189.000
NorthEast	11.200	110.0	3.400	1029.300	1729.200	516.700
NorthEast	16.300	159.0	6.300	1244.500	1873.400	615.000
NorthEast	26.100	254.0	11.100	1348.200	2312.300	1109.600
South	9.300	81.0	5.700	415.800	824.900	107.100
South	16.900	180.0	8.725	891.000	1265.950	213.100
South	20.950	223.5	12.850	985.950	1462.900	297.050
South	25.575	264.0	14.650	1200.575	2030.050	357.350
South	31.900	337.0	17.400	1857.200	3048.600	547.800
West	14.200	46.0	2.600	699.700	2237.800	207.000
West	20.200	120.0	4.900	989.300	2614.300	301.200
West	29.300	161.0	6.800	1535.500	2865.500	408.200
West	38.700	263.0	9.000	1607.700	2988.500	545.600
West	46.000	294.0	12.200	2149.800	3720.100	635.900

Note:

Here are general comments of the table.

[1] First value of region for column is at 0%;

[2] Second value of region for column is at 25%

[a] Third value of region for column is at 50%;

[b] Fourth value of region for column is at 75%;

[*] Fifth value of region for column is at 100%

```
USArrestscombinedImportantCorrelation <- USArrestscombinedImportant %>%  select_if(is.numeric) %>% select(contain
s('Rates'))
USArrestscombinedImportantCorrelation$Murder <- USArrestscombinedImportant$Murder
USArrestscombinedImportantCorrelation$Assault <- USArrestscombinedImportant$Assault
USArrestscombinedImportantCorrelation$Rape <- USArrestscombinedImportant$Rape
USArrestscombinedImportantCorrelation <- USArrestscombinedImportantCorrelation%>% select(-c(Data.Rates.Property.A
ll, Data.Rates.Violent.All)) %>% rename(
    Burglary = Data.Rates.Property.Burglary,
    Larceny = Data.Rates.Property.Larceny,
    Motor = Data.Rates.Property.Motor,
    Robbery = Data.Rates.Violent.Robbery,
    )
```

```
library("Hmisc")
corMatrix <- rcorr(as.matrix(USArrestscombinedImportantCorrelation))
as.data.frame(corMatrix$r) %>% kbl(caption = "Correlation Matrix of Different State Crimes in 1973") %>%
  kable_classic(full_width = F, html_font = "Cambria") %>% kable_material_dark()
```

Correlation Matrix of Different State Crimes in 1973

	Burglary	Larceny	Motor	Robbery	Murder	Assault	Rape
Burglary	1.0000000	0.7800244	0.6437819	0.6149926	0.3331508	0.5621404	0.7787774
Larceny	0.7800244	1.0000000	0.4897432	0.3459301	-0.0333869	0.3225882	0.6489864
Motor	0.6437819	0.4897432	1.0000000	0.6415893	0.0808595	0.3431231	0.4194696
Robbery	0.6149926	0.3459301	0.6415893	1.0000000	0.4921382	0.5804301	0.6056911
Murder	0.3331508	-0.0333869	0.0808595	0.4921382	1.0000000	0.8018733	0.5635788
Assault	0.5621404	0.3225882	0.3431231	0.5804301	0.8018733	1.0000000	0.6652412
Rape	0.7787774	0.6489864	0.4194696	0.6056911	0.5635788	0.6652412	1.0000000

```
cor(USArrestscombinedImportantCorrelation) %>%
  # Save as a data frame
  as.data.frame %>%
  # Convert row names to an explicit variable
  rownames_to_column %>%
  # Pivot so that all correlations appear in the same column
  pivot_longer(-1, names_to = "other_var", values_to = "correlation") %>%
  # Specify variables are displayed alphabetically from top to bottom
  ggplot(aes(rowname, factor(other_var, levels = rev(levels(factor(other_var)))), fill=correlation)) +
  # Heatmap with geom_tile
  geom_tile() +
  # Change the scale to make the middle appear neutral
  scale_fill_gradient2(low="red",mid="white",high="blue") +
  # Overlay values
  geom_text(aes(label = round(correlation,2)), color = "black", size = 4) +
  # Give title and labels
  labs(title = "Correlation matrix for State Crime Rates in 1973", x = "", y = "")
```

Correlation matrix for State Crime Rates in 1973