Assignment 3

Instructor: Prof. Gunnar Heins

Teaching Assistant: Pedro de Sousa Almeida

Student: C.H.(Chenhui Lu)

UFID: 76982846

==========

```
# Empty Environment
rm(list = ls())
```

```
#library
library("data.table")
```

```
## Warning: package 'data.table' was built under R version 4.3.3
```

```
library("stargazer")
```

```
##
## Please cite as:
```

```
##  Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics
Tables.
```

```
##  R package version 5.2.3. https://CRAN.R-project.org/package=stargazer
```

```
library("AER")
```

```
## Warning: package 'AER' was built under R version 4.3.2
```

```
## Loading required package: car
```

```
## Loading required package: carData
```

```
## Loading required package: lmtest
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:data.table':
##
##     yearmon, yearqtr
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
## Loading required package: sandwich
```

```
## Loading required package: survival
```

```
## Warning: package 'survival' was built under R version 4.3.2
```

```
library("ggplot2")
```

```
## Warning: package 'ggplot2' was built under R version 4.3.2
```

```
library("ggrepel")
```

```
## Warning: package 'ggrepel' was built under R version 4.3.3
```

```
library("maps")
library("dplyr")
```

```
##
## Attaching package: 'dplyr'
```

```
## The following object is masked from 'package:car':
##
##     recode
```

```
## The following objects are masked from 'package:data.table':
##
##     between, first, last
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library("mapproj")
```

1 Local Labor Markets [35 points]

Frequently researchers use variation across regions for empirical research. Especially useful is that the U.S. census collects a wide range of data on U.S. counties. There are more than 3,000 such counties and for each of them you can and detailed information on demographics, the economy, employment, or elections. Often researchers use this type of data to draw conclusions on the people that live in this areas. For example, if a county has a very educated population on average as well as a high income per capita, one might conclude that education translates into higher earnings. In this question, I ask you to check if some common trends can be identified from county data. Variables: fipstate: Code, identifying a state (e.g. 1 for Alabama, 2 for Arizona, etc) fipscty: Code, identifying a county within a state (Alachua County in Florida has code 1, for example) share white i: Share of white individuals in year i share black i: Share of black individuals in year i share college i: Share of individuals with at least a college degree in year i share female i: Share of women in a county in year i income i: Income per capita in year i mfg i: Number of workers employed in manufacturing in year i population i: County population in year i

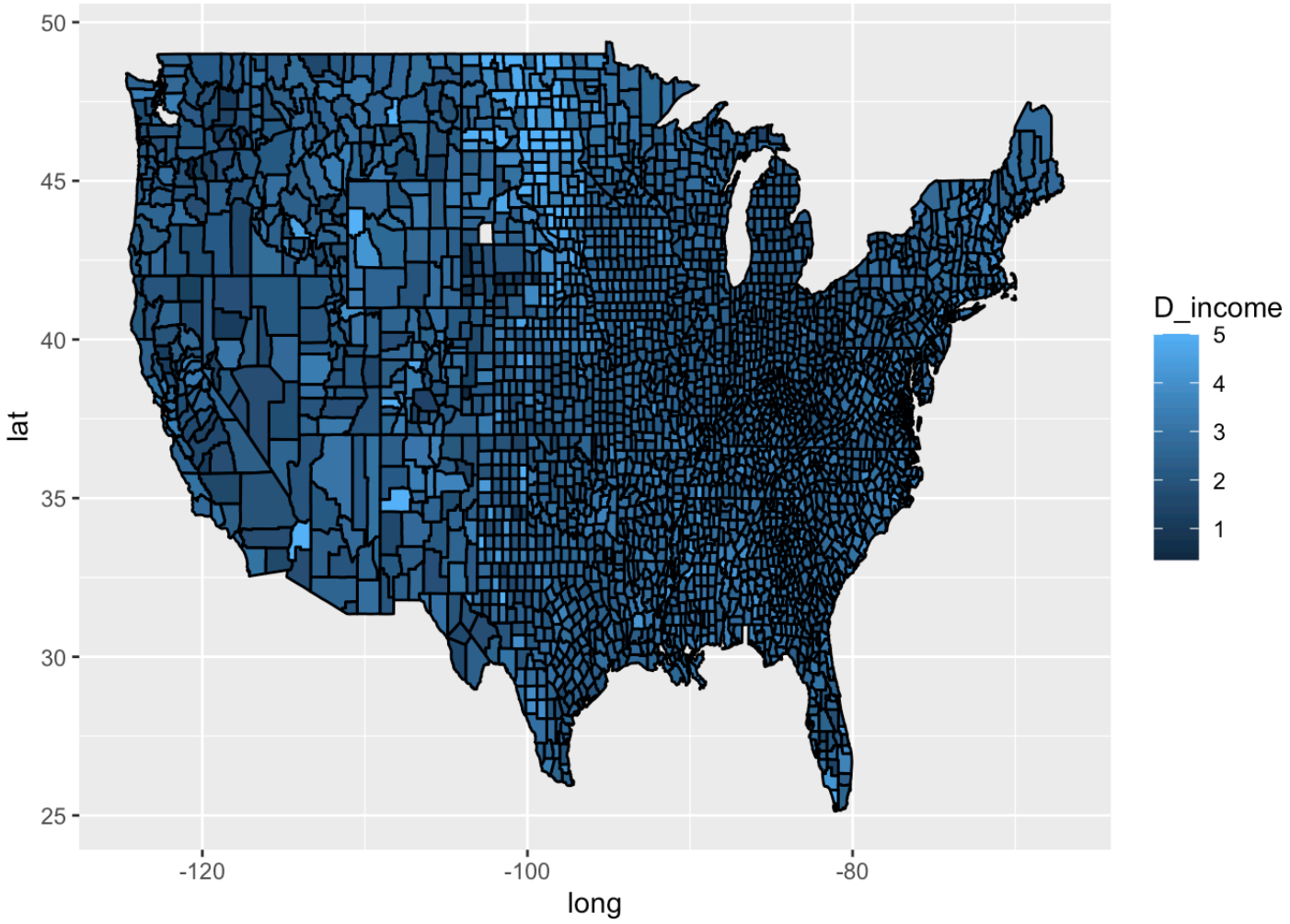
Questions:

1. Import the dataset CountyDemographics.csv using fread(). Create a variable with the percentage change in income between 1990 and 2010 and add it to the dataset. Delete observations for which this percentage change is NA or Inf.

```
# inport the data
dt <- fread("/Users/terrylu/Desktop/UF/fall/R and Matlab/Matlab/assignment/Problem Se
t 5/CountyDemographics.csv")

dt <- dt[, income_change_1090 := (income_10 - income_90)/income_90]    #create the in
come change percentage
dt <- dt[is.na(income_change_1090) == FALSE]                           # delete the o
bservations with income_change of N/A
dt <- dt[is.infinite(income_change_1090) == FALSE]                     # delete the o
bservations with income_change of Inf
```

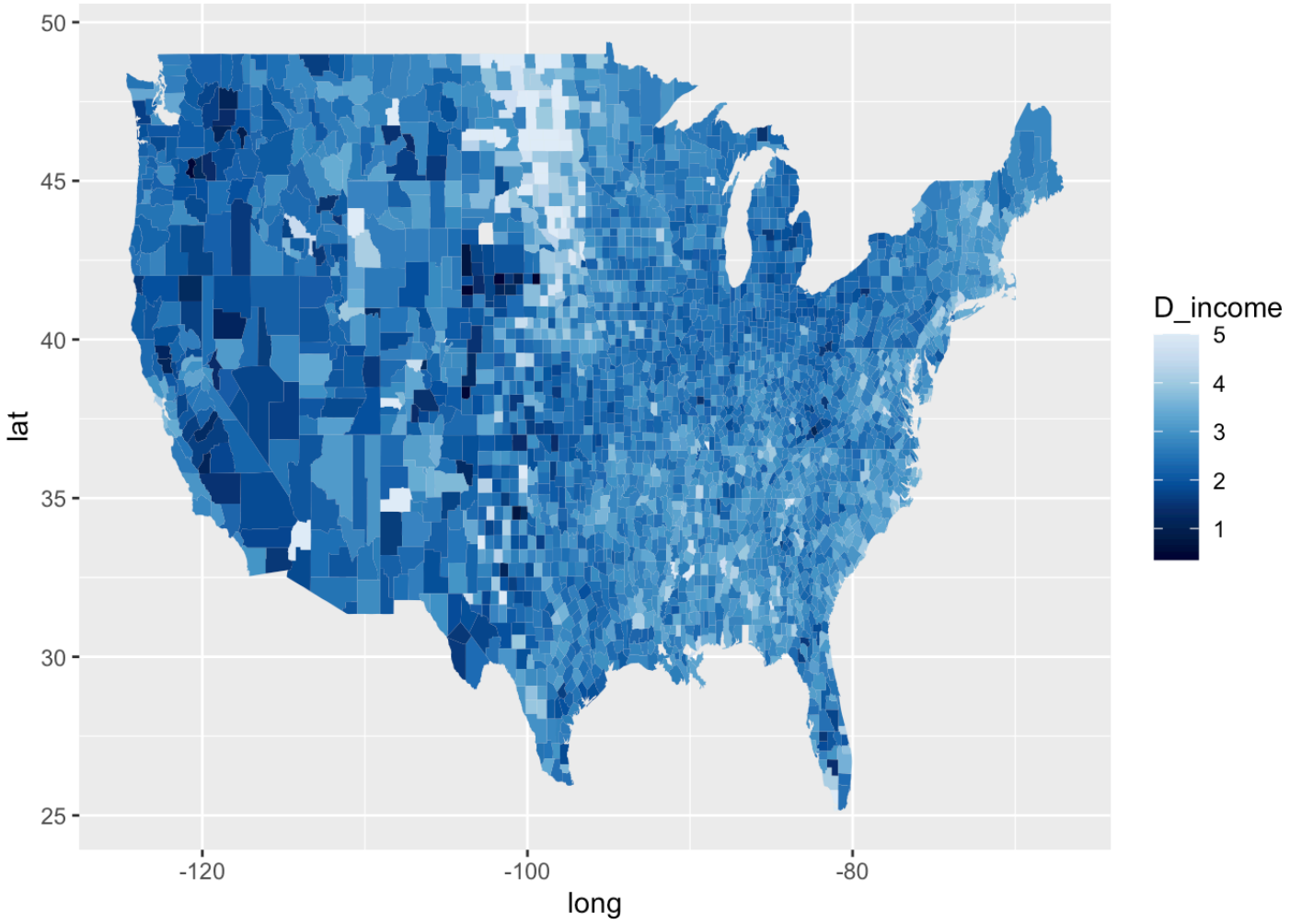
2. The following commands create a map in R. In order to do so, first install packages “maps”, “dplyr” and “mapproj” and load them via library(“maps”), library(“dplyr”) and library(“mapproj”). What does this particular map that R creates show?

```
DataCounty <- map_data("county")
CountyDemographics <- fread("/Users/terrylu/Desktop/UF/fall/R and Matlab/Matlab/assig
nment/Problem Set 5/CountyDemographics.csv")

DataCounty <- merge(DataCounty, CountyDemographics, by = c("region", "subregion"), al
l = F)   #merge datasets by "region", "subregion"

DataCounty <- as.data.table(DataCounty)
setkey(DataCounty, order)

DataCounty[ , D_income := (income_10 - income_80) / income_80]      #generate the per
centage income change between 1910 and 1980
DataCounty[D_income > 5 , D_income := 5]                            #any value of D_i
ncome bigger 5, changes to 5
ggplot(data = DataCounty, aes(x = long, y = lat, fill = D_income, group = group)) + g
eom_polygon(colour="black")                                  #crate the map by D_change.
```



I prefer no border color for polygons. And color from black to blue to white.

```
ggplot(data = DataCounty, aes(x = long, y = lat, fill = D_income, group = group)) + g
eom_polygon() + scale_fill_gradientn(colors = c("#000033", "#041f4b", "#08306b", "#08
519c", "#2171b5", "#4292c6", "#6baed6", "#9ecae1", "#c6dbef", "#deebf7"))
```



This heatmap shows different percentage income change in different county. Brighter area shows higher persentage income change.

3. During the last presidential campaigns, it has often been pointed out that especially white men with comparably low education have suffered economically. Regress the percentage change in income between 1980 and 2010 on the share of white residents in a county in 1990, the share of women in 1990, and the share of college educated residents in 1990. What do you find?

```
dt <- dt[ , D_income_1080 := (income_10 - income_80) / income_80]      #generate the
percentage income change between 1910 and 1980
dt <- dt[is.na(D_income_1080) == FALSE]                                # delete the observ
ations with income_change of N/A
dt <- dt[is.infinite(D_income_1080) == FALSE]                          # delete the observ
ations with income_change of Inf
reg_3 <- lm(D_income_1080 ~ share_white_90 + female_90 + college_90, data = dt)    #r
egession
summary(reg_3)
```

```
##
## Call:
## lm(formula = D_income_1080 ~ share_white_90 + female_90 + college_90,
##     data = dt)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.3744 -0.3928 -0.0903  0.2650 12.5668
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     2.69526    0.43227   6.235 5.13e-10 ***
## share_white_90 -0.38200    0.09399  -4.064 4.94e-05 ***
## female_90       0.47203    0.89830   0.525    0.599
## college_90      1.40771    0.22031   6.390 1.92e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7816 on 3061 degrees of freedom
## Multiple R-squared:  0.01824,    Adjusted R-squared:  0.01727
## F-statistic: 18.95 on 3 and 3061 DF,  p-value: 3.541e-12
```

1. The share of white residents (share_white_90) has a negative significant impact on percentage income change. For each 1% increase in the share of white residents (coefficient = -0.382), the percentage change in income will decrease by 0.382%.
2. The share of women (female_90) doesn't has a significant impact on percentage income change.
3. The share of college-educated residents (college_90) has a positive significant impact on percentage income change. For each 1% increase in the share of college-educated residents (coefficient = 1.40771), the percentage change in income will increase by 1.40771%.

4. Another common claim is that international trade has resulted in high employment losses since the 1990s. As particularly manufacturing goods are traded, researchers often use the manufacturing employment share as a measure of how exposed a region or country is exposed to trade. In order to evaluate this claim, let's import an additional dataset on manufacturing employment: Match the dataset data inc pop.csv to CountyDemographics using merge and the variable "fips" as identifier.

```
inc_pop <- fread("/Users/terrylu/Desktop/UF/fall/R and Matlab/Matlab/assignment/Probl
em Set 5/data_inc_pop.csv")
dtmg <- merge(inc_pop, dt, by = "fips")
```

5. Create the manufacturing employment share by dividing the number of workers in manufacturing by the population in each county. Create this variable for the years 1990, 2000, and 2010.

```
dtmg <- dtmg[, emp_share_90 := mfg_90/population_90]   #Create the manufacturing empl
oyment share
dtmg <- dtmg[, emp_share_00 := mfg_00/population_00]
dtmg <- dtmg[, emp_share_10 := mfg_10/population_10]
```

6. Compute the average manufacturing employment share in 1990, 2000, and 2010 over all counties. How did it change over time?

```
dtmg <- dtmg[, ave_share_90 := mean(emp_share_90)]  #Compute the average manufacturin
g employment share
dtmg <- dtmg[, ave_share_00 := mean(emp_share_00)]
dtmg <- dtmg[, ave_share_10 := mean(emp_share_10)]
dtmg_1 <- dtmg[, .SD[1]]

dtmg_1$ave_share_90
```

```
## [1] 0.06044608
```

```
dtmg_1$ave_share_00
```

```
## [1] 0.05800853
```

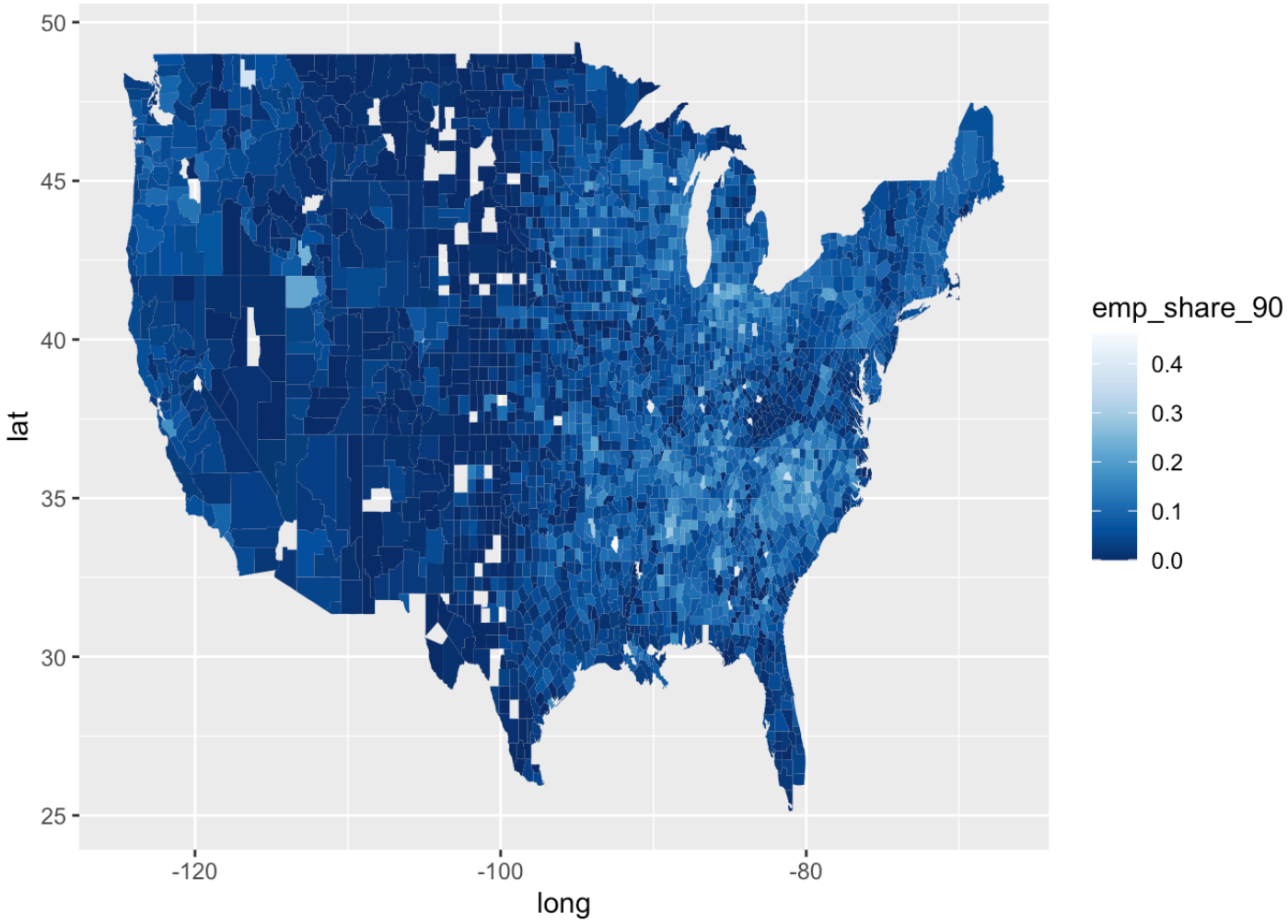
```
dtmg_1$ave_share_10
```

```
## [1] 0.03746714
```

The average manufacturing employment share is decreasing with the time.

7. Create a map similarly to the one created in part 2 that shows counties according to their manufacturing employment share in 1990. Do you see much overlap with the map you created in part (2)?

```
map_dtmg <- merge(DataCounty,dtmg, by = c("region", "subregion"), all = F)   #merge d
atasets by "region", "subregion"

ggplot(data = map_dtmg, aes(x = long, y = lat, fill = emp_share_90, group = group)) +
geom_polygon() + scale_fill_gradientn(colors = c("#08306b", "#08519c", "#2171b5", "#4
292c6", "#6baed6", "#9ecae1", "#c6dbef", "#deebf7","#f7fbff"))
```




```
# create the heatmap, I prefer no border color for polygons. Color from black to blue to white.
```

No, the two heatmaps do not overlap, actually they are opposite to some extent.

8. Regress the percentage change in income on the manufacturing share in 1990. Include the regressors from (3) as controls. What do you find?

First of all, I don't understand the meaning of percentage change in income in 1990. If we want to talk about the change, there should be 2 time points, however, here's only one time point (1990). So I assume that the problem want to ask the change between 1980 to 2010 to fit the question (3).

```
reg8 <- lm(income_change_1090 ~ emp_share_90 + share_white_90 + female_90 + college_90, data = dtmg)
summary(reg8)
```

```
##
## Call:
## lm(formula = income_change_1090 ~ emp_share_90 + share_white_90 +
##     female_90 + college_90, data = dtmg)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.00138 -0.12825 -0.01005  0.11996  1.64556
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     0.64382    0.14044   4.584 4.74e-06 ***
## emp_share_90   -0.78537    0.08190  -9.589  < 2e-16 ***
## share_white_90 -0.12787    0.02904  -4.403 1.11e-05 ***
## female_90       0.93156    0.28873   3.226  0.00127 **
## college_90      0.36035    0.06747   5.341 9.96e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2374 on 3003 degrees of freedom
## Multiple R-squared:  0.05597,    Adjusted R-squared:  0.05471
## F-statistic: 44.51 on 4 and 3003 DF,  p-value: < 2.2e-16
```

The coefficient of the manufacturing share in 1990 (emp_share_90) has a significantly negative impact on the the percentage change in income on actuaally, which is coincide with what maps show.

```
%%%%%%%%%%%%%%%%%%%%
%   Assignment 5   %
% Apart II and III %
%%%%%%%%%%%%%%%%%%%%

clear               % Empty the Workspace
clc                 % Clear the Screen
format compact      % compact format

%%%2 Loops in Matlab (40 points)

%%1. Compute
sum1 = 0;                           %initial sum = 0
for i = 1:100                       %set up the loop
    sum1 = sum1 + 1/sqrt(i);
end
sum1                                %output
% the sum = 18.5896

%%2. Compute
sum2 = 0;                               %initial sum = 0
for i = 1:20                            %set up the double loops
    for j = 1 : 20
        if i > j
            sum2 = sum2 + (i - j)^2;
        end
    end
end
sum2                                     %output
% sum2 = 13300


%%3. Compute
e = 0;                                   %initial sum = 0
for n = 1:100                            %set up the loop
    e = e + (1/factorial(n));
end
e                                        %output
% e = 1.7183


%%4. Create the following 10-by-10 matrix
M = 5 * ones(10,10) ;                    %build up a 10*10 metrix with 5
for i = 1:length(M);                     %set up double loops
    for j = 1:length(M)
        if  i == j;
            M(i,j) = 0;                  %make the elements in principal diagonal = 0
        end
    end
end
M                                        %output


%%5. Create two matrices Y and X and compute
X = randi(10,20,40);                        % creat the random (1,20) metrix X:20*40
Y = randi(10,20,1);                         % creat the random (1,20) metrix Y:20*1
beta = (X' * X) \ (X' * Y);                 %calculate the beta
epsilon = Y - X*beta;                       %calculate the epsilon
```

```
beta
epsilon


%%%3 Functions in Matlab (25 points)

%%1. Write a function that takes two matrices, A and B, as input and returns 2 arguments:
%(1) y = 1 if matrix A has more rows than matrix B and y = 0 otherwise
%(2) z = 1 if matrix A has more columns than matrix B and z = 0 otherwise
[Y,Z] = compare(X,Y)
%Please check the compare.m

%%2. Suppose you have two vectors, x and y, each with 3 distinct elements.
%Write a function that returns the number of elements that x and y have in common
%(e.g. if the vectors are c(1,2,3) and c(3,6,2), it should return 2).

%Please check the number_com.m


%%3. Write a function that returns the largest value of an n n matrix
% if n is an even number and the minimum otherwise.
%Please check the maxormin.m
A = randi(20,40,40);
z = maxormin(A);
z
```

```
%%%Function 3.1

function [Y,Z] = compare(A,B)              % set up the function
    if  size(A,1) > size(B,1)              % if number of rows of A is bigger
        Y = 1;
    else
        Y = 0;
    end
    if  size(A,2) > size(B,2)              % if number of columns of A is bigger
        Z = 1;
    else
        Z = 0;
    end
end
```

```
%%%function 3.2

function [z] = number_com(x,y)              % set up the function
z = 0                                       %initial z = 0
x_flat = x(:);                              %not necessary in this case: make the merix to↙
be a vector
y_flat = y(:);
for i = 1:length(x_flat)                    %build up double loops
    for j = 1:length(y_flat)
        if x_flat(i) == y_flat(j)
            z = z + 1;                      %count the number of same elements
        end
    end
end
end
```

```
%%%Function 3.3

function [z] = maxormin(m)              % set up the function
x = size(m,1);                          % only need the number of rows
if mod(x, 2) == 0                       % if x is even
    z = max(m(:));                      % z is the maximum
else
    z = min(m(:));                      % otherwise z si the minimum
end
end
```