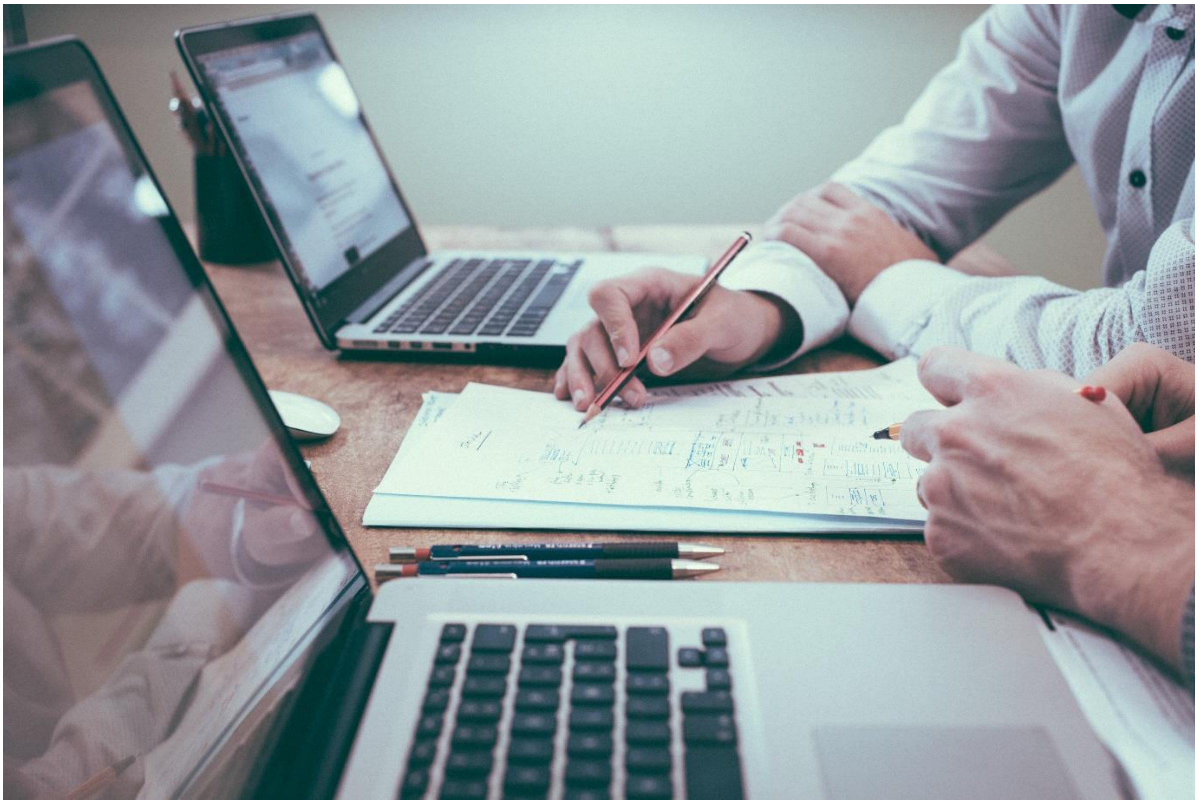


Statistical analysis of the EU data in R Female Inactivity Rate, aged 25-54 Dataset LFST_R_LFP2ACTRTN

Franklin Chukwuma Ofoegbu

Course: Statistical analysis of the EU data in R @SGGW

1/22/23

Statistical Analysis of the EU Data - **lfst_r_lfp2actrtn**

Analysis for Females in the EU aged 25-54, inactivity rate

1. Inactivity Rate: is the percentage of inactive persons in relation to the comparable total population .Inactive persons meaning persons part of the Labour force but are not employed/working.
   Inactivity Rate = 100 - Activity Rate
   (Activity rate is the percentage of active persons in relation to the comparable total population.)
   Economic Importance: In an economy or in the EU, we can know in each age group who are economically active/inactive. It will help for planning.
2. Data is downloaded from EUROSTAT

```
# Step 2 - Download the data (directly from Eurostat, using eurostat library)
activity_rate_eu <- get_eurostat("lfst_r_lfp2actrtn", time_format = "num")
ls(activity_rate_eu)
unique(activity_rate_eu$unit)
unique(activity_rate_eu$geo)
unique(activity_rate_eu$isced11)
unique(activity_rate_eu$sex)
unique(activity_rate_eu$time)
unique(activity_rate_eu$citizen)
unique(activity_rate_eu$age)
```

   The unique values are also seen.
3. Data is cleaned and filtered.

```
activity_rate_eu <- activity_rate_eu %>%
  mutate(    edu =
               case_when(
                 isced11 == "ED0-2" ~ "edu_low",
                 isced11 == "ED3_4" ~ "edu_medium",
                 isced11 == "ED5-8" ~ "edu_high",
                 isced11 == "NRP" ~ "NRP",
                 isced11 == "TOTAL" ~ "edu_total"
               )
  ) %>%
  select(geo, unit, age, sex, edu, time, values, citizen)
```

   We have to obtain the inactivity rate, which is

```
# The inactivity rate is gotten as
inactivity_rate_eu <- activity_rate_eu
inactivity_rate_eu$values <- 100 - inactivity_rate_eu$values
```

   First 5 rows of the activity rate in eu data

geo <chr>	unit <chr>	age <chr>	sex <chr>	edu <chr>	time <dbl>	values <dbl>	citizen <chr>
AT	PC	Y15-64	F	edu_low	2021	56.5	EU27_2020_FOR
AT1	PC	Y15-64	F	edu_low	2021	55.0	EU27_2020_FOR
AT11	PC	Y15-64	F	edu_low	2021	NA	EU27_2020_FOR
AT12	PC	Y15-64	F	edu_low	2021	59.6	EU27_2020_FOR
AT13	PC	Y15-64	F	edu_low	2021	53.3	EU27_2020_FOR

5 rows

   First 5 rows of the inactivity rate in eu data

geo <chr>	unit <chr>	age <chr>	sex <chr>	edu <chr>	time <dbl>	values <dbl>	citizen <chr>
AT	PC	Y15-64	F	edu_low	2021	43.5	EU27_2020_FOR
AT1	PC	Y15-64	F	edu_low	2021	45.0	EU27_2020_FOR
AT11	PC	Y15-64	F	edu_low	2021	NA	EU27_2020_FOR
AT12	PC	Y15-64	F	edu_low	2021	40.4	EU27_2020_FOR
AT13	PC	Y15-64	F	edu_low	2021	46.7	EU27_2020_FOR

5 rows

In AT geo, Activity Rate + Inactivity Rate = 56.5+ 43.5 = 100
Which verifies our data.

Moving on to filtering the Regions for NUTS 2 and NUTS 0

The Values of NUTS 0 are seen below:

geo <chr>
AT
BE
BG
CY
CZ
DE
DK
EE
EL
ES

1-10 of 27 rows

The Values of NUTS 2 are seen below:

geo <chr>
AT11
AT12
AT13
AT21
AT22
AT31
AT32
AT33
AT34
BE10

1-10 of 235 rows

A Snapshot for inactivity rate of females aged 25-54 in NUTS 0, 2019 are shown below

```
# NUTS 0
# inactivity rate 2019 Females 25-54 NUTS 0
inactivity_rate_eu_2019_F_0 <- inner_join(inactivity_rate_eu, eu_region_0) %>%  #merge the preliminary data frame with selected regions
  filter(age == "Y25-54") %>%
  filter(sex == "F") %>%
  filter(time== 2019) %>%
  filter(edu!="edu_total") %>%
  filter(edu!="NRP") %>%
  filter(citizen=="TOTAL") %>%
  select(geo, edu, time, values)
```

A Snapshot for inactivity rate of females aged 25-54 in NUTS 2, 2019 are shown below

```
# NUTS 2
# inactivity rate 2019 Females 25-54 NUTS 2
inactivity_rate_eu_2019_F_2 <- inner_join(inactivity_rate_eu, eu_region_2) %>%  #merge the preliminary data frame with selected regions
  filter(age == "Y25-54") %>%
  filter(sex == "F") %>%
  filter(time== 2019) %>%
  filter(edu!="edu_total") %>%
  filter(edu!="NRP") %>%
  filter(citizen=="TOTAL") %>%
  select(geo, edu, time, values)
```

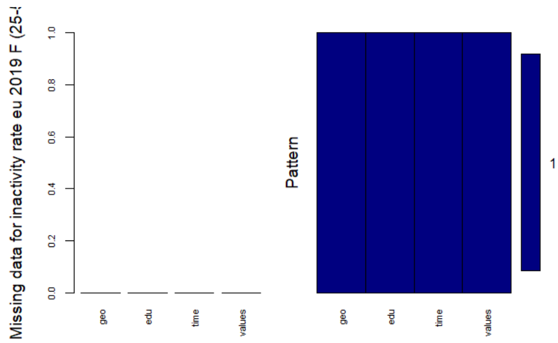
This implies that the data to be analyzed is divided into six categories:

1. 2019 EU Inactivity Rate for Females Aged 25-54 in NUTS 0 which has label inactivity_rate_eu_2019_F_0
2. 2020 EU Inactivity Rate for Females Aged 25-54 in NUTS 0 which has label inactivity_rate_eu_2020_F_0
3. 2021 EU Inactivity Rate for Females Aged 25-54 in NUTS 0 which has label inactivity_rate_eu_2021_F_0
4. 2019 EU Inactivity Rate for Females Aged 25-54 in NUTS 2 which has label inactivity_rate_eu_2019_F_2
5. 2020 EU Inactivity Rate for Females Aged 25-54 in NUTS 2 which has label inactivity_rate_eu_2020_F_2
6. 2021 EU Inactivity Rate for Females Aged 25-54 in NUTS 2 which has label inactivity_rate_eu_2021_F_2

STEP 4: Preliminary Analysis and Interpretation

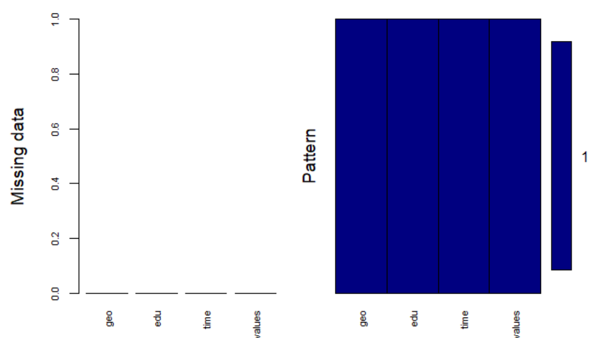
Identifying the missing data:

- inactivity_rate_eu_2019_F_0



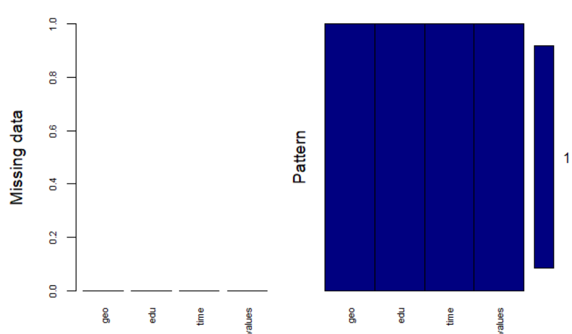
No missing data here

- inactivity_rate_eu_2020_F_0



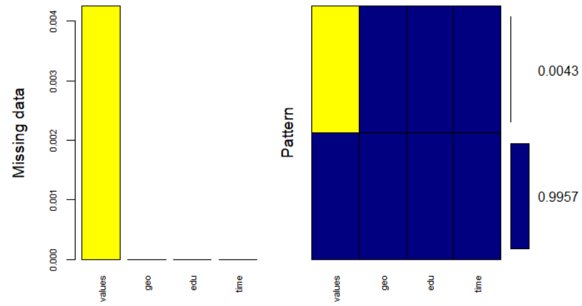
No missing data

- inactivity_rate_eu_2021_F_0



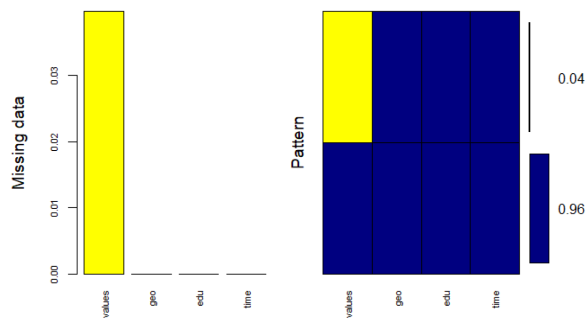
No missing data

- inactivity_rate_eu_2019_F_2



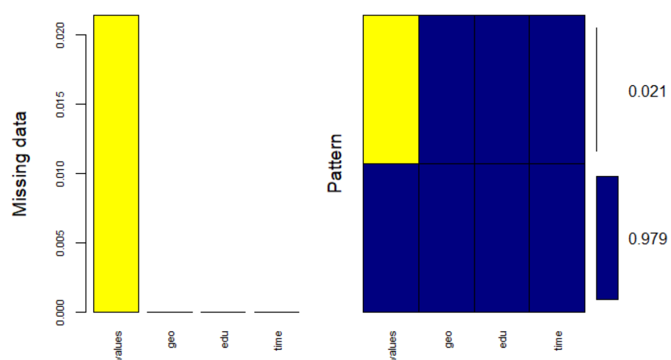
0.0043% of data are missing.

- inactivity_rate_eu_2020_F_2



0.04% of data are missing.

- inactivity_rate_eu_2021_F_2



0.021% of data are missing.

Some data in the NUTS 2 inactivity rate are missing.

4b. Imputing missing values in the datasets using the MICE method.
The datasets with missing values are imputed using the MICE PMM (Predictive Matching Method) algorithm.

Careful observation of the datasets shows that there are no more missing values after computation.

```
> summary(inactivity_rate_eu_2019_F_2)
     geo                 edu                 time          values     
 Length:705         Length:705         Min.   :2019   Min.   : 3.30  
 Class :character   Class :character   1st Qu.:2019   1st Qu.:11.20  
 Mode  :character   Mode  :character   Median :2019   Median :18.00  
                                       Mean   :2019   Mean   :23.29  
                                       3rd Qu.:2019   3rd Qu.:33.20  
                                       Max.   :2019   Max.   :76.00  
```

```
> summary(inactivity_rate_eu_2020_F_2)
     geo                 edu                 time          values     
 Length:705         Length:705         Min.   :2020   Min.   : 1.70  
 Class :character   Class :character   1st Qu.:2020   1st Qu.:11.90  
 Mode  :character   Mode  :character   Median :2020   Median :18.70  
                                       Mean   :2020   Mean   :23.14  
                                       3rd Qu.:2020   3rd Qu.:32.20  
                                       Max.   :2020   Max.   :72.70  
> summary(inactivity_rate_eu_2021_F_2)
     geo                 edu                 time          values     
 Length:702         Length:702         Min.   :2021   Min.   : 3.70  
 Class :character   Class :character   1st Qu.:2021   1st Qu.:10.82  
 Mode  :character   Mode  :character   Median :2021   Median :17.50  
                                       Mean   :2021   Mean   :23.22  
                                       3rd Qu.:2021   3rd Qu.:33.50  
                                       Max.   :2021   Max.   :81.10  
```

4c. Identifying Outliers in the datasets

There are no outliers in the NUTS 0 Data Set, but there are outliers in the NUTS 2 Data Set

inactivity_rate_eu_2019_F_2

geo <chr>	edu <chr>	time <dbl>	values <dbl>	is.outlier <lgl>	is.extreme <lgl>
FRM0	edu_low	2019	70.7	TRUE	FALSE
ITF3	edu_low	2019	68.9	TRUE	FALSE
ITF4	edu_low	2019	66.9	TRUE	FALSE
ITF6	edu_low	2019	67.3	TRUE	FALSE
ITC1	edu_low	2019	71.3	TRUE	FALSE
PL82	edu_low	2019	66.4	TRUE	FALSE
RO12	edu_low	2019	76.0	TRUE	FALSE

Outliers for inactivity rates in EU NUTS 2, 2019 Females Aged 25-54

inactivity_rate_eu_2020_F_2

geo <chr>	edu <chr>	time <dbl>	values <dbl>	is.outlier <lgl>	is.extreme <lgl>
BE32	edu_low	2020	65.5	TRUE	FALSE
BG31	edu_low	2020	62.8	TRUE	FALSE
ITF3	edu_low	2020	72.6	TRUE	FALSE
ITF4	edu_low	2020	68.7	TRUE	FALSE
ITF5	edu_low	2020	67.5	TRUE	FALSE
ITF6	edu_low	2020	69.1	TRUE	FALSE
ITC1	edu_low	2020	72.7	TRUE	FALSE
RO12	edu_low	2020	70.3	TRUE	FALSE

Outliers for inactivity rates in EU NUTS 2, 2020 Females Aged 25-54

inactivity_rate_eu_2021_F_2

geo	edu	time	values	is.outlier	is.extreme
<chr>	<chr>	<dbl>	<dbl>	<lgl>	<lgl>
DED5	edu_low	2021	81.1	TRUE	FALSE
ITF3	edu_low	2021	71.2	TRUE	FALSE
ITF4	edu_low	2021	68.0	TRUE	FALSE
ITF6	edu_low	2021	68.6	TRUE	FALSE
ITG1	edu_low	2021	74.2	TRUE	FALSE
PL91	edu_low	2021	81.1	TRUE	FALSE
RO12	edu_low	2021	81.1	TRUE	FALSE
RO22	edu_low	2021	70.3	TRUE	FALSE
RO31	edu_low	2021	67.7	TRUE	FALSE
RO42	edu_low	2021	70.9	TRUE	FALSE

1-10 of 12 rows Previous 1 2 Next

Outliers for inactivity rates in EU NUTS 2, 2021 Females Aged 25-54

It is seen that the highest inactivity rates come from ITF 3, ITF 4 and RO12. The inactivity rate in Italy is supported by this article.

4d. Calculating the mean, standard deviation, minimum value, max value and range of the data

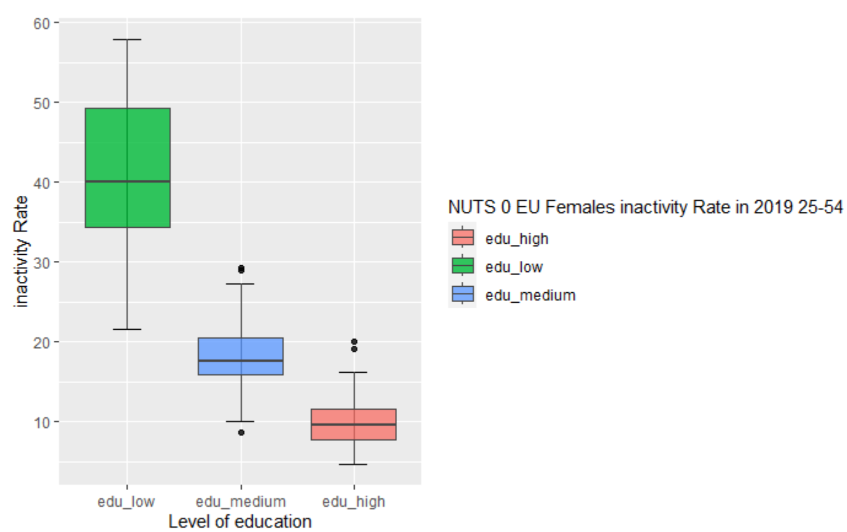
	mean	Standard Dev	min	max	range least	range end
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
2019 nuts 0	23.15556	14.46391	4.7	57.9	4.7	57.9
2020 nuts 0	23.73086	14.74241	4.0	58.8	4.0	58.8
2021 nuts 0	23.13827	15.40090	4.6	64.9	4.6	64.9
2019 nuts 2	23.29106	14.93556	3.3	76.0	3.3	76.0
2020 nuts 2	23.13730	14.26995	1.7	72.7	1.7	72.7
2021 nuts 2	23.22222	15.79422	3.7	81.1	3.7	81.1

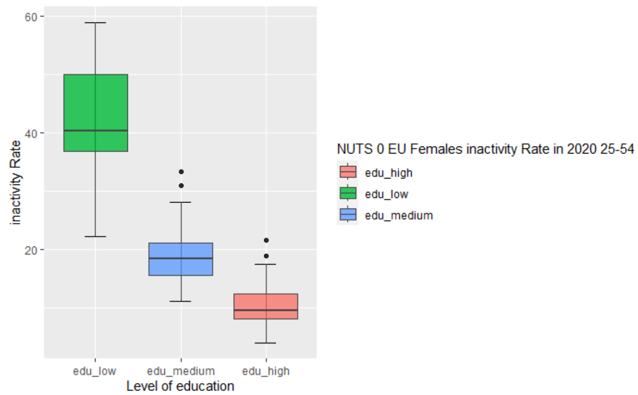
6 rows

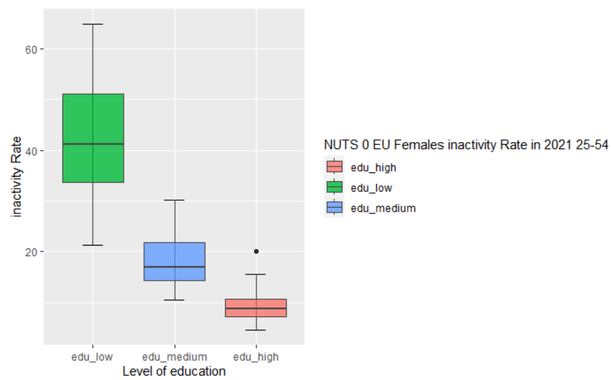
The max inactivity rates and largest deviations occurred in 2021 for both NUTS 0 and 2. This might be the aftermath of COVID on the economy.

4e. Boxplots

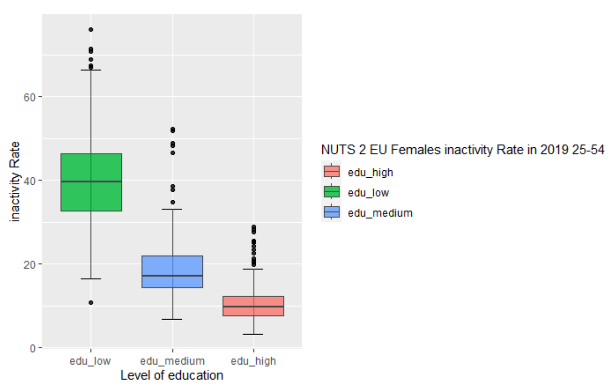
NUTS 0

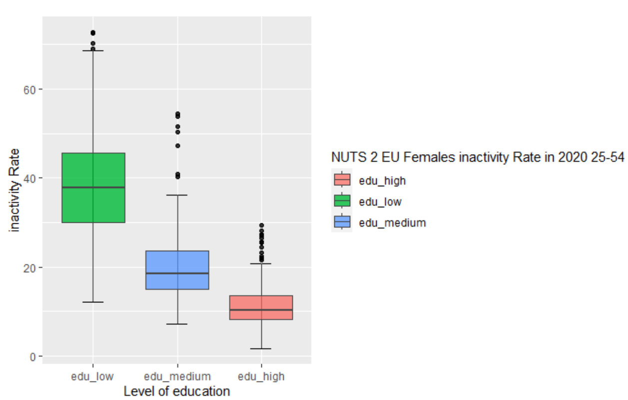


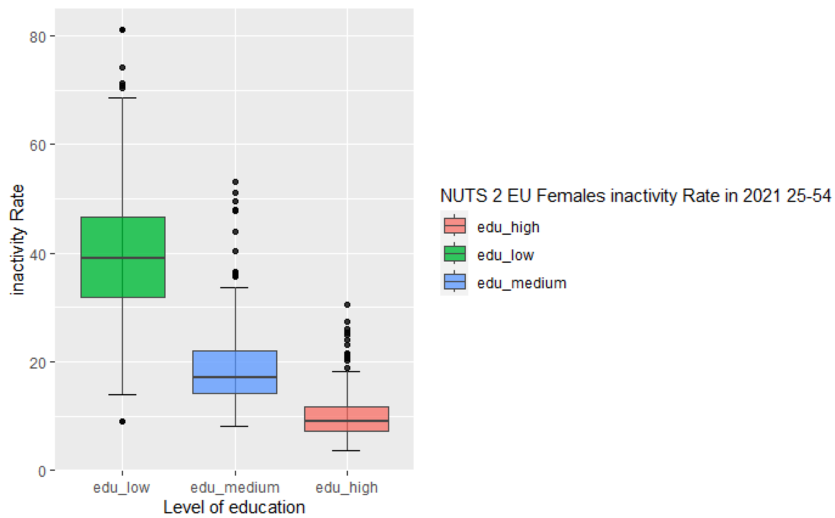

60
40
20
inactivity Rate
edu_low
edu_medium
edu_high
Level of education
NUTS 0 EU Females inactivity Rate in 2020 25-54
edu_high
edu_low
edu_medium


60
40
20
inactivity Rate
edu_low
edu_medium
edu_high
Level of education
NUTS 0 EU Females inactivity Rate in 2021 25-54
edu_high
edu_low
edu_medium

NUTS 2


60
40
20
0
inactivity Rate
edu_low
edu_medium
edu_high
Level of education
NUTS 2 EU Females inactivity Rate in 2019 25-54
edu_high
edu_low
edu_medium


60
40
20
0
inactivity Rate
edu_low
edu_medium
edu_high
Level of education
NUTS 2 EU Females inactivity Rate in 2020 25-54
edu_high
edu_low
edu_medium

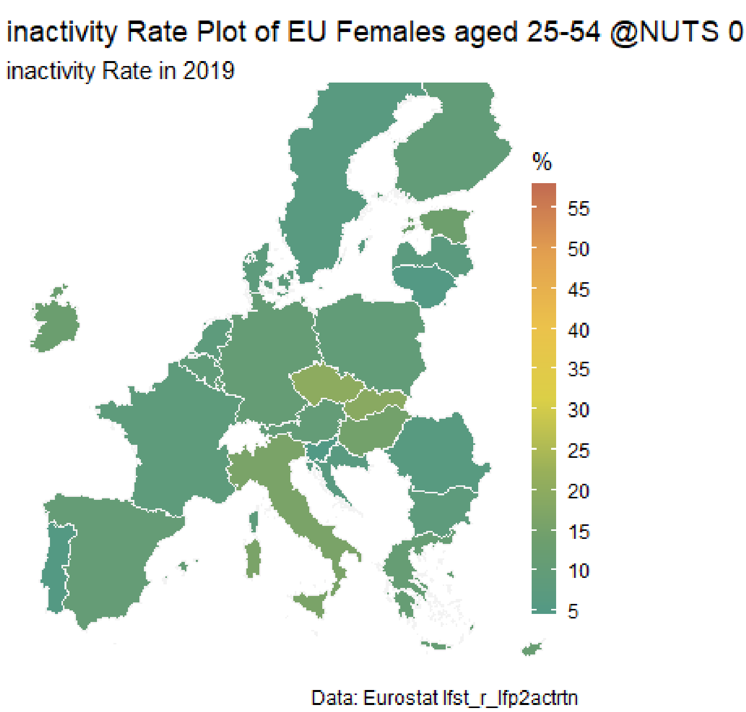


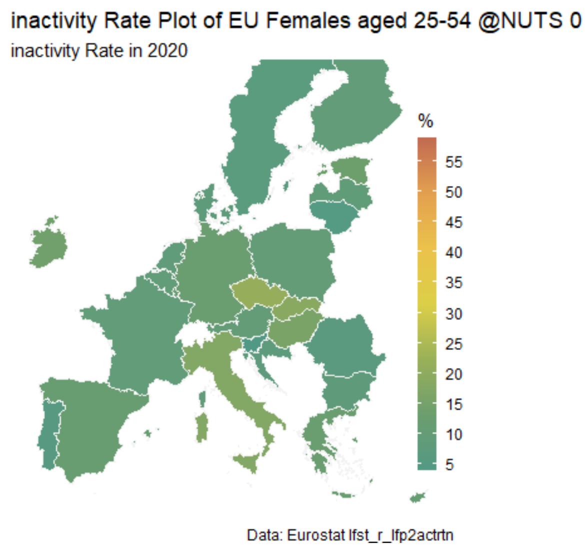
The outliers were not removed, due to the fact that they are not extreme outliers and that they might represent some key info on the dataset.

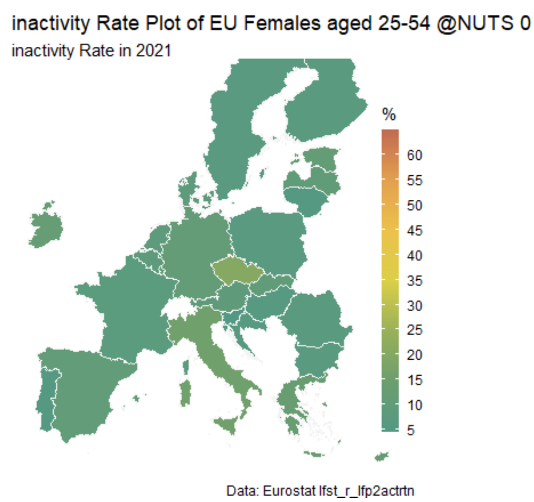
Hence, the higher the education level, the lower the inactivity rate seen.

4f. Maps

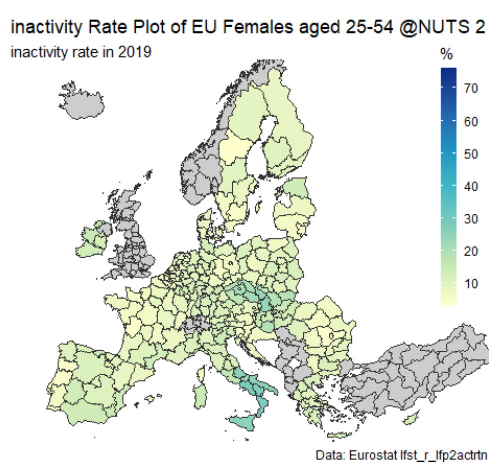
NUTS 0 level.

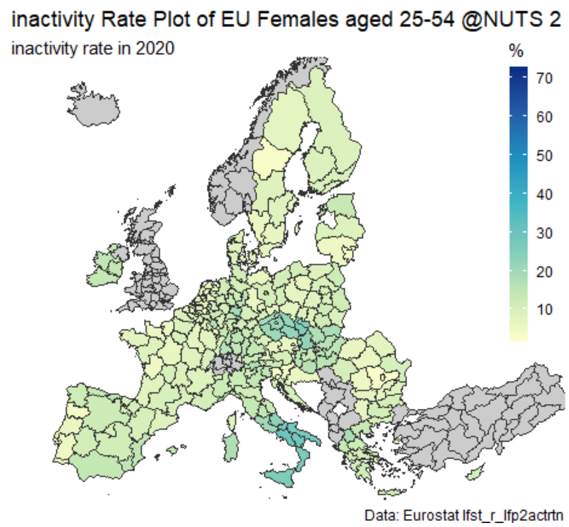


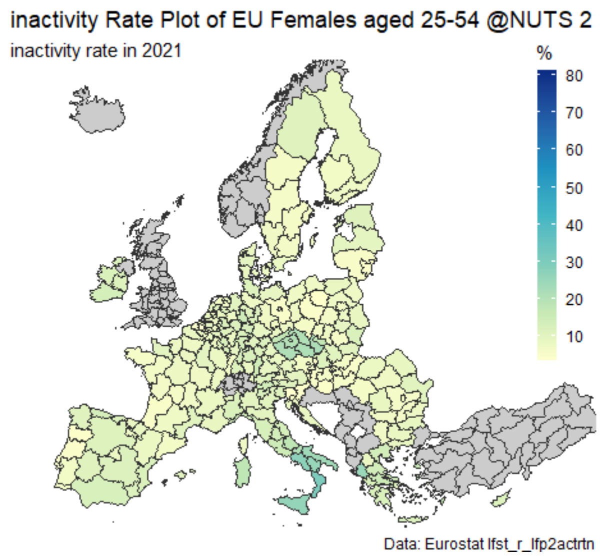

inactivity Rate Plot of EU Females aged 25-54 @NUTS 0
inactivity Rate in 2020
%
55
50
45
40
35
30
25
20
15
10
5
Data: Eurostat lfst_r_lfp2actrtn


inactivity Rate Plot of EU Females aged 25-54 @NUTS 0
inactivity Rate in 2021
%
60
55
50
45
40
35
30
25
20
15
10
5
Data: Eurostat lfst_r_lfp2actrtn

NUTS 2


inactivity Rate Plot of EU Females aged 25-54 @NUTS 2
inactivity rate in 2019
%
70
60
50
40
30
20
10
Data: Eurostat lfst_r_lfp2actrtn





6. ANOVA Analysis

a. Across Educational Level (edu vs inactivity)
   The Shapiro-Wilk Normality test will be used to test for Normality in the datasets.

   $\alpha = 0.05$

   H0: The distribution of the dataset is normal.
   H1: The distribution of the dataset is not normal.

```
        Shapiro-Wilk normality test

data:  inactivity_rate_eu_2019_F_0$values
W = 0.90254, p-value = 1.397e-05


        Shapiro-Wilk normality test

data:  inactivity_rate_eu_2020_F_0$values
W = 0.9062, p-value = 2.015e-05


        Shapiro-Wilk normality test

data:  inactivity_rate_eu_2021_F_0$values
W = 0.89466, p-value = 6.5e-06


        Shapiro-Wilk normality test

data:  inactivity_rate_eu_2019_F_2$values
W = 0.90552, p-value < 2.2e-16


        Shapiro-Wilk normality test

data:  inactivity_rate_eu_2020_F_2$values
W = 0.90929, p-value < 2.2e-16


        Shapiro-Wilk normality test

data:  inactivity_rate_eu_2021_F_2$values
W = 0.88867, p-value < 2.2e-16
```

Since the p-values < 0.05, H0 cannot be accepted.

Shapiro-Wilk Results

Dataset	Shapiro-Wilk Results	Conclusion
inactivity_rate_eu_2019_F_0	W = 0.90254, p-value = 1.397e-05	H0 cannot be accepted; dataset is not normal
inactivity_rate_eu_2020_F_0	W = 0.9062, p-value = 2.015e-05	H0 cannot be accepted; dataset is not normal
inactivity_rate_eu_2021_F_0	W = 0.89466, p-value = 6.5e-06	H0 cannot be accepted; dataset is not normal
inactivity_rate_eu_2019_F_2	W = 0.90552, p-value < 2.2e-16	H0 cannot be accepted; dataset is not normal
inactivity_rate_eu_2020_F_2	W = 0.90929, p-value < 2.2e-16	H0 cannot be accepted; dataset is not normal
inactivity_rate_eu_2021_F_2	W = 0.88867, p-value < 2.2e-16	H0 cannot be accepted; dataset is not normal

Hence Non-Parametric ANOVA Test will be used. The Kruskal-Wallis Test is employed.

Using the Kruskal-Wallis Rank sum Test
H0: All 3 educational levels have the same distribution (The Educational Levels does not affect the inactivity Rate)
H1: Not all 3 educational levels have the same distribution. (The Educational Level affect the inactivity Rate)

Taking α = 0.05, the results are seen below

```
        Kruskal-Wallis rank sum test

data:  values by edu
Kruskal-Wallis chi-squared = 63.341, df = 2, p-value = 1.761e-14


        Kruskal-Wallis rank sum test

data:  values by edu
Kruskal-Wallis chi-squared = 63.218, df = 2, p-value = 1.872e-14


        Kruskal-Wallis rank sum test

data:  values by edu
Kruskal-Wallis chi-squared = 66.32, df = 2, p-value = 3.97e-15


        Kruskal-Wallis rank sum test

data:  values by edu
Kruskal-Wallis chi-squared = 515.8, df = 2, p-value < 2.2e-16


        Kruskal-Wallis rank sum test

data:  values by edu
Kruskal-Wallis chi-squared = 459.79, df = 2, p-value < 2.2e-16


        Kruskal-Wallis rank sum test

data:  values by edu
Kruskal-Wallis chi-squared = 459.79, df = 2, p-value < 2.2e-16
```

Since the p-values < 0.05, H0 cannot be accepted.

Kruskal-Wallis Results

Dataset	Kruskal-Wallis Results	Conclusion
inactivity_rate_eu_2019_F_0	X^2= 63.341, df = 2, p-value = 1.761e-14	H0 cannot be accepted; Posthoc Tests
inactivity_rate_eu_2020_F_0	X^2= 63.218, df = 2, p-value = 1.872e-14	H0 cannot be accepted; Posthoc Tests
inactivity_rate_eu_2021_F_0	X^2= 66.32, df = 2, p-value = 3.97e-15	H0 cannot be accepted; Posthoc Tests
inactivity_rate_eu_2019_F_2	X^2= 515.8, df = 2, p-value < 2.2e-16	H0 cannot be accepted; Posthoc Tests
inactivity_rate_eu_2020_F_2	X^2= 459.79, df = 2, p-value < 2.2e-16	H0 cannot be accepted; Posthoc Tests
inactivity_rate_eu_2021_F_2	X^2= 459.79, df = 2, p-value < 2.2e-16	H0 cannot be accepted; Posthoc Tests

Hence, not all 3 educational levels have the same distribution. (The Educational Level affect the inactivity Rate)

Post-hoc Tests – Dunn's Test

p-values adjusted with the Bonferroni method.

inactivity_rate_eu_2019_F_0

Comparison <chr>	Z <dbl>	P.unadj <dbl>	P.adj <dbl>
edu_high - edu_low	-7.930336	2.185543e-15	6.556630e-15
edu_high - edu_medium	-3.383841	7.147932e-04	2.144379e-03
edu_low - edu_medium	4.546494	5.454682e-06	1.636405e-05

inactivity_rate_eu_2020_F_0

Comparison <chr>	Z <dbl>	P.unadj <dbl>	P.adj <dbl>
edu_high - edu_low	-7.930470	2.183180e-15	6.549541e-15
edu_high - edu_medium	-3.470665	5.191709e-04	1.557513e-03
edu_low - edu_medium	4.459805	8.203433e-06	2.461030e-05

inactivity_rate_eu_2021_F_0

Comparison <chr>	Z <dbl>	P.unadj <dbl>	P.adj <dbl>
edu_high - edu_low	-8.141602	3.900813e-16	1.170244e-15
edu_high - edu_medium	-3.910283	9.218812e-05	2.765644e-04
edu_low - edu_medium	4.231319	2.323246e-05	6.969737e-05

inactivity_rate_eu_2019_F_2

Comparison <chr>	Z <dbl>	P.unadj <dbl>	P.adj <dbl>
edu_high - edu_low	-22.651052	1.362212e-113	4.086635e-113
edu_high - edu_medium	-9.893882	4.425286e-23	1.327586e-22
edu_low - edu_medium	12.757170	2.843452e-37	8.530355e-37

inactivity_rate_eu_2020_F_2

Comparison <chr>	Z <dbl>	P.unadj <dbl>	P.adj <dbl>
edu_high - edu_low	-21.43937	5.738963e-102	1.721689e-101
edu_high - edu_medium	-10.39422	2.634435e-25	7.903306e-25
edu_low - edu_medium	11.04515	2.313806e-28	6.941419e-28

inactivity_rate_eu_2021_F_2

Comparison <chr>	Z <dbl>	P.unadj <dbl>	P.adj <dbl>
edu_high - edu_low	-21.43937	5.738963e-102	1.721689e-101
edu_high - edu_medium	-10.39422	2.634435e-25	7.903306e-25
edu_low - edu_medium	11.04515	2.313806e-28	6.941419e-28

DunnTest performs the post hoc pairwise multiple comparisons procedure appropriate to follow up a Kruskal-Wallis test, which is a non-parametric analog of the one-way ANOVA.

Observing carefully, the adjusted and unadjusted p-values, taking $\alpha = 0.05$; all the values are insignificant.

Hence, all the educational values edu-high, edu-medium and edu-low all independently affect the inactivity rate.

b. Across the years
   Data description:

1. EU Inactivity Rate for Females Aged 25-54 in NUTS 0 with high education which has label inactivity_rate_eu_F_0_high.
2. EU Inactivity Rate for Females Aged 25-54 in NUTS 0 with medium education which has label inactivity_rate_eu_F_0_medium.
3. EU Inactivity Rate for Females Aged 25-54 in NUTS 0 with low education which has label inactivity_rate_eu_F_0_low.

4. EU Inactivity Rate for Females Aged 25-54 in NUTS 2 with high education which has label inactivity_rate_eu_F_2_high.
5. EU Inactivity Rate for Females Aged 25-54 in NUTS 2 with medium education which has label inactivity_rate_eu_F_2_medium.
6. EU Inactivity Rate for Females Aged 25-54 in NUTS 2 with low education which has label inactivity_rate_eu_F_2_low.

$\alpha = 0.05$

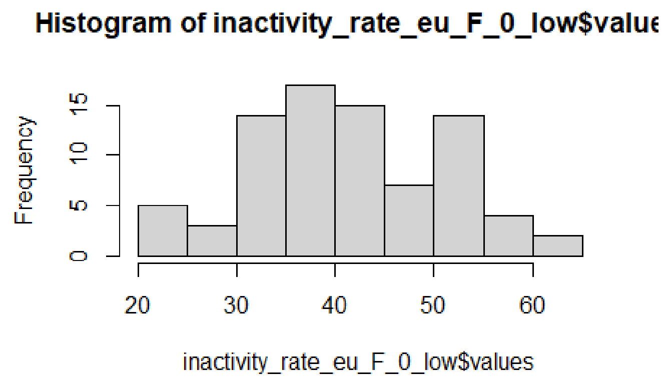
H0: The distribution of the dataset is normal.
H1: The distribution of the dataset is not normal.

Performing the Normality Tests on the data

```
        Shapiro-Wilk normality test

data:  inactivity_rate_eu_F_0_high$values
W = 0.92366, p-value = 0.000129


        Shapiro-Wilk normality test

data:  inactivity_rate_eu_F_0_medium$values
W = 0.94084, p-value = 0.0009845


        Shapiro-Wilk normality test

data:  inactivity_rate_eu_F_0_low$values
W = 0.98306, p-value = 0.3618


        Shapiro-Wilk normality test

data:  inactivity_rate_eu_F_2_high$values
W = 0.88066, p-value < 2.2e-16


        Shapiro-Wilk normality test

data:  inactivity_rate_eu_F_2_medium$values
W = 0.8755, p-value < 2.2e-16


        Shapiro-Wilk normality test

data:  inactivity_rate_eu_F_2_low$values
W = 0.98198, p-value = 1.286e-07
```

Shapiro-Wilk Results

Dataset	Shapiro-Wilk Results	Conclusion
inactivity_rate_eu_F_0_high	W = 0.92366, p-value = 0.000129	H0 cannot be accepted; dataset is not normal
inactivity_rate_eu_F_0_medium	W = 0.94084, p-value = 0.0009845	H0 cannot be accepted; dataset is not normal
inactivity_rate_eu_F_0_low	W = 0.98306, p-value = 0.3618	H0 cannot be rejected; dataset is not normal
inactivity_rate_eu_F_2_high	W = 0.88066, p-value < 2.2e-16	H0 cannot be accepted; dataset is not normal
inactivity_rate_eu_F_2_medium	W = 0.8755, p-value < 2.2e-16	H0 cannot be accepted; dataset is not normal
inactivity_rate_eu_F_2_low	W = 0.98198, p-value = 1.286e-07	H0 cannot be accepted; dataset is not normal



For low education in NUTS 0, the data seems Normal, following the Shapiro-Wilk Test.

Using the Kruskal-Wallis Rank sum Test for the datasets that are not Normal.
H0: All 3 time/year levels have the same distribution (The time Levels does not affect the inactivity Rate)
H1: Not all 3 time/year levels have the same distribution. (The time Level affect the inactivity Rate)

Taking α = 0.05, the results are seen below

```
        Kruskal-Wallis rank sum test

data:  values by time
Kruskal-Wallis chi-squared = 1.6114, df = 2, p-value = 0.4468


        Kruskal-Wallis rank sum test

data:  values by time
Kruskal-Wallis chi-squared = 0.45087, df = 2, p-value = 0.7982


        Kruskal-Wallis rank sum test

data:  values by time
Kruskal-Wallis chi-squared = 0.21245, df = 2, p-value = 0.8992


        Kruskal-Wallis rank sum test

data:  values by time
Kruskal-Wallis chi-squared = 13.743, df = 2, p-value = 0.001037


        Kruskal-Wallis rank sum test

data:  values by time
Kruskal-Wallis chi-squared = 5.2534, df = 2, p-value = 0.07232


        Kruskal-Wallis rank sum test

data:  values by time
Kruskal-Wallis chi-squared = 3.4275, df = 2, p-value = 0.1802
```

Dataset	Kruskal-Wallis Results	Conclusion
inactivity_rate_eu_F_0_high	X^2= 1.6114, df = 2, p-value = 0.4468	H0 cannot be rejected
inactivity_rate_eu_F_0_medium	X^2= 0.45087, df = 2, p-value = 0.7982	H0 cannot be rejected
inactivity_rate_eu_F_0_low	X^2= 0.21245, df = 2, p-value = 0.8992	Already follows normal distribution
inactivity_rate_eu_F_2_high	X^2= 13.743, df = 2, p-value = 0.001037	H0 cannot be accepted; Posthoc Tests
inactivity_rate_eu_F_2_medium	X^2= 5.2534, df = 2, p-value = 0.07232	H0 cannot be rejected
inactivity_rate_eu_F_2_low	X^2= 3.4275, df = 2, p-value = 0.1802	H0 cannot be rejected

Further tests could still be carried out, but the relationship between the year/time and the inactivity rate can be concluded by not rejecting H0.
This means that for most of the groups of data used in the analysis, the effect of years on the inactivity rate is minimal/nonexistent. The years 2019, 2020 and 2021 have the same distribution on the inactivity rate.

7. Conclusion

The educational levels edu-high, edu-medium and edu-low all independently affect the inactivity rate (do not have the same distributions).
The years/periods observed: 2019, 2020 and 2021 almost all have the same distribution on the activity rate (there is no significant difference in their effects on EU Female Activity Rate).
The more educated an EU female in the age range 25 – 54, the lower their inactivity rate.
MICE PMM algorithm is an effective tool for generating missing values in a dataset, and was employed here.
It is seen that the highest inactivity rates come from ITF 3, ITF 4 and RO12. The inactivity rate in Italy is supported by this article.