IST5128 - Take Home Final Exam (Due Date: 16/01/2023, 13:00)

- PLEASE READ THE LAST PAGE OF THESE QUESTIONS TO LEARN ABOUT OTHER DETAILS ABOUT TAKE-HOME FINAL EXAM
- This take home final exam consists of two parts.
  - Part A is about creating an artificial dataset and making some analysis based on it.
  - Part B is about a water quality data set. You will use your data manipulation skills to create a clean and tidy dataset and data visualization skills to draw some graphs.

PART A: Create an Artificial Dataset (Total 30 Points)

Let's create an artificial dataset with size **100000** which consists of five variables namely `age`, `team`, `rating`, `income` and `grade`.

The definition for these variables is given below:

- `age`: Age of the students. Takes values between 18 and 25. The probability of all values are equal. This variable is independent from all other variables.
- `team`: Favorite team of the students. Can take values `Fenerbahce`, `Galatasaray` and `Besiktas` with 0.25 probability each, and can also take `Trabzonspor` with 0.10 probability and `Other` with 0.15 probability. This variable is independent from all other variables.
- `rating`: Interest of a student to a particular course. 1 is the least, 5 is the most. This variable is independent from all other variables. The probability of all values are equal.
- `income`: Takes values between 3000TL to 10000TL with increments of 100TL $(3000, 3100, \ldots, 9900, 10000)$. This variable is independent from all other variables. The probability of all values are equal.
- `grade`: Takes values between 1 and 100. This variable depends on the favorite team and rating. There are four different rules given in the table.

Team and Rating	[1-30]	[31-40]	[41-60]	[61-90]	[91-100]
Fenerbahce and rating is 4 or 5	0	0.01	0.3	0.6	0.09
Fenerbahce and rating is 1,2 or 3	0.3	0.2	0.2	0.3	0
Not Fenerbahce and rating is 4 or 5	0.3	0.1	0.2	0.3	0.1
Not Fenerbahce and rating is 1,2 or 3	0.6	0.1	0.3	0	0

Q1 (5 PTS)

- Create `Age`, `Team`, `Rating` and `Income` with given specifications.
- You can do this by using the `sample()` function with suitable arguments.
- Combine variables `Age`, `Team`, `Rating` and `Income` to a dataframe called `Q1`.
- You should get similar mean values and frequencies given below.

```
> table(age)

age
   18    19    20    21    22    23    24    25
12247 12710 12539 12362 12606 12522 12509 12505

> mean(age)

[1] 21.50497

> table(team)

team
   Besiktas  Fenerbahce Galatasaray       Other Trabzonspor
      25235       25074       24735       14972        9984

> prop.table(table(team))

team
   Besiktas  Fenerbahce Galatasaray       Other Trabzonspor
    0.25235     0.25074     0.24735     0.14972     0.09984

> table(rating)

rating
    1     2     3     4     5
20011 20104 19971 19993 19921

> mean(rating)

[1] 2.99709

> mean(income)

[1] 6495.011
```

The dataset rows might be different due to randomness.

```
> head(Q1)

  age        team rating income
1  19    Besiktas      2   5200
2  24    Besiktas      4   9900
3  23 Galatasaray      4   3300
4  25  Fenerbahce      5   8400
5  20    Besiktas      2   9900
6  18    Besiktas      2   3600
```

Q2) (10 PTS)

- Now create the `grade` variable.
- There may be multiple ways to create the grade variable, but one way could be using `mutate()` and `case_when()` with `sample()` function by defining the probabilities for each situation.
- Add the newly created `grade` variable to the dataframe `Q1` and call this dataframe `Q2`

The dataset rows might be different due to randomness.

```
> head(Q2,15)
```

	age	team	rating	income	grade
1	19	Besiktas	2	5200	35
2	24	Besiktas	4	9900	71
3	23	Galatasaray	4	3300	64
4	25	Fenerbahce	5	8400	52
5	20	Besiktas	2	9900	32
6	18	Besiktas	2	3600	24
7	19	Other	3	3600	3
8	23	Besiktas	3	8400	15
9	21	Galatasaray	4	7200	42
10	24	Fenerbahce	5	9000	77
11	25	Galatasaray	4	4100	13
12	23	Other	3	8000	58
13	23	Besiktas	4	4700	95
14	21	Besiktas	5	5400	73
15	24	Other	4	3100	64

Q3) (5 PTS)

- Find the mean grade based on each team and rating.
- Do you see the pattern? Make comment on your finding.

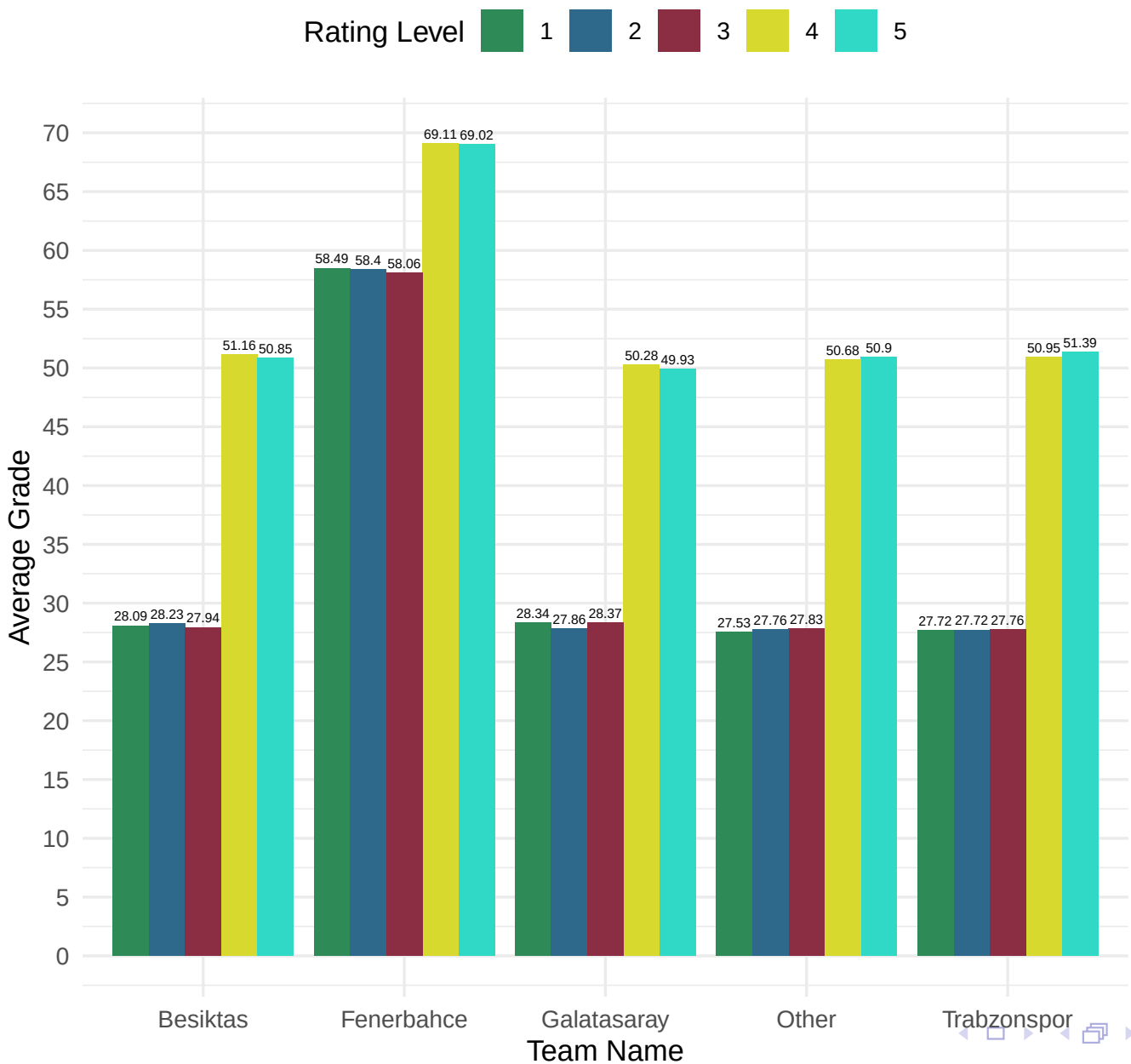
The dataset rows might be different due to randomness but you should get similar mean grades.

```
> Q3 %>% print(n=Inf)
# A tibble: 25 × 3
# Groups:   team [5]
   team        rating mean_grade
   <chr>        <int>      <dbl>
 1 Fenerbahce       4       69.1
 2 Fenerbahce       5       69.0
 3 Fenerbahce       1       58.5
 4 Fenerbahce       2       58.4
 5 Fenerbahce       3       58.1
 6 Trabzonspor      5       51.4
 7 Besiktas         4       51.2
 8 Trabzonspor      4       50.9
 9 Other            5       50.9
10 Besiktas         5       50.8
11 Other            4       50.7
12 Galatasaray      4       50.3
13 Galatasaray      5       49.9
14 Galatasaray      3       28.4
15 Galatasaray      1       28.3
16 Besiktas         2       28.2
17 Besiktas         1       28.1
18 Besiktas         3       27.9
19 Galatasaray      2       27.9
20 Other            3       27.8
21 Other            2       27.8
22 Trabzonspor      3       27.8
23 Trabzonspor      2       27.7
24 Trabzonspor      1       27.7
25 Other            1       27.5
```

Q4) (10 PTS)

- Create the following graph based on the output of `Q3`
- The hex codes for the bars are `#2E8B57`,`#2E688B`, `#8B2E43`,`#D8D92F` and `#2FD9C6`.
- You should use `geom_text` to add average values to the graph. Search internet for more details.

Q4) GRAPH


Average grade based on Team Name and Rating
Rating Level
1
2
3
4
5
Average Grade
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
28.09
28.23
27.94
51.16
50.85
58.49
58.4
58.06
69.11
69.02
28.34
27.86
28.37
50.28
49.93
27.53
27.76
27.83
50.68
50.9
27.72
27.72
27.76
50.95
51.39
Besiktas
Fenerbahce
Galatasaray
Other
Trabzonspor
Team Name

PART B: Water Quality Data (Total 70 Points)

In the second part you will tidy an untidy water quality dataset and then you will make some visualization from the tidy data. The data is given in `data.csv` and it contains the following columns.

- `X`: Station Number and Station Name, 11 unique values
- `X.1`: Sub Header for parameter: 59 Unique Parameter values
- `Parameter`: 59 Unique Parameter values
- `Unit`: Measurement Unit for the parameter
- `Year`: 1985 to 2013
- `X1-X12`: Month where X1 is January and X12 is the December

Import the dataset, with the additional argument `na.strings`.

Tidied version of the dataset is also provided to you in the `final_data.xlsx` to understand the aim of the question.

PART B: Water Quality Data - First Look at the Data

```
> head(data,10)
```

	X	X.1	Parameter
1	Station No: 01-11-00-001; Station Name: MERIC NEHRI-KAPIKULE	<NA>	<NA>
2	<NA>	Al	<NA>
3	<NA>	<NA>	Al
4	<NA>	As	<NA>
5	<NA>	<NA>	As
6	<NA>	<NA>	As
7	<NA>	<NA>	As
8	<NA>	<NA>	As
9	<NA>	BOD5	<NA>
10	<NA>	<NA>	BOD5

	Unit	Year	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
1	<NA>	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
2	<NA>	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
3	µg/L	2013	NA	NA	NA	168.25	NA	156.87	NA	NA	NA	161.21	NA	NA
4	<NA>	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
5	µg/L	1987	NA	NA	NA	4.00	NA	1.00	NA	NA	NA	6.00	NA	NA
6	µg/L	1992	NA	NA	32	NA	NA	NA	NA	NA	NA	NA	NA	NA
7	µg/L	1994	NA	NA	NA	NA	NA	9.00	NA	NA	NA	NA	NA	NA
8	µg/L	2013	NA	3.49	NA	1.17	NA	52.20	NA	NA	NA	2.74	NA	NA
9	<NA>	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
10	mg/L	1985	NA	NA	NA	5.10	NA	6.10	NA	7.2	NA	7.70	NA	7.3

PART B: Water Quality Data - First Look at the Data

There are few problems in the initial dataset.

- The months are given in columns with `X1` to `X12` and not with a single column.
- The variable `X` defines the station name as a header but it only appears in the first line of the station name.Following rows only have `NA` values until the next station name.
- Since `X` is a header, if it is not NA, then all the other columns are NA
- It seems that `X.1` is the header for parameters and always gets NA values at the other columns when `X.1` is not NA.
- `Parameter` also has the same problem as in the column `X`.

Q5 (10 PTS)

- Start by completing the NA values in the rows after the station name.
- You can do this by using the `fill()` function.
- Then, keep only NA values in the column `X.1` by combining `filter()` and `is.na()` functions.
- As a third step, keep only not NA values in the column `Parameter` with `filter()` and `is.na()` functions.
- Second and third step, would remove unnecessary empty rows, arose from the headers.
- As a last step, rename the variable `X` as `Station_Name`

```
> head(Q5,10)
```

	Station_Name	Parameter	Year	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
1	Station No: 01-11-00-001; Station Name: MERIC NEHRI-KAPIKULE	Al	2013	NA	NA	NA	168.25	NA	156.87	NA	NA	NA	161.21	NA	NA
2	Station No: 01-11-00-001; Station Name: MERIC NEHRI-KAPIKULE	As	1987	NA	NA	NA	4.00	NA	1.00	NA	NA	NA	6.00	NA	NA
3	Station No: 01-11-00-001; Station Name: MERIC NEHRI-KAPIKULE	As	1992	NA	NA	32.0	NA	NA	NA	NA	NA	NA	NA	NA	NA
4	Station No: 01-11-00-001; Station Name: MERIC NEHRI-KAPIKULE	As	1994	NA	NA	NA	NA	NA	9.00	NA	NA	NA	NA	NA	NA
5	Station No: 01-11-00-001; Station Name: MERIC NEHRI-KAPIKULE	As	2013	NA	3.49	NA	1.17	NA	52.20	NA	NA	NA	2.74	NA	NA
6	Station No: 01-11-00-001; Station Name: MERIC NEHRI-KAPIKULE	BOD5	1985	NA	NA	NA	5.10	NA	6.10	NA	7.2	NA	7.70	NA	7.3
7	Station No: 01-11-00-001; Station Name: MERIC NEHRI-KAPIKULE	BOD5	1986	NA	NA	6.5	7.00	NA	4.90	NA	3.6	NA	6.80	NA	9.7
8	Station No: 01-11-00-001; Station Name: MERIC NEHRI-KAPIKULE	BOD5	1987	NA	9.20	NA	3.80	NA	4.20	NA	NA	NA	3.80	NA	9.6
9	Station No: 01-11-00-001; Station Name: MERIC NEHRI-KAPIKULE	BOD5	1988	NA	9.30	NA	2.70	NA	5.70	NA	7.0	NA	2.10	NA	6.7
10	Station No: 01-11-00-001; Station Name: MERIC NEHRI-KAPIKULE	BOD5	1989	NA	5.00	NA	3.60	NA	6.00	NA	6.7	NA	4.50	NA	8.0

Q6 (5 PTS)

- If you look at the structure of `Q5` you will see that `Year` is in the numeric format.
- Convert this to a year, month,day format with `ymd()` function in the `lubridate` package.
- You should define the `truncated` argument in the `ymd()` function as `2L` to parse undefined month and day.
- Since day and month is unknown, parsed date will be in "Year-01-01" format.

```
> head(Q6,10)
```

	Station_Name	Parameter
1	Station No: 01-11-00-001; Station Name: MERIC NEHRI-KAPIKULE	Al
2	Station No: 01-11-00-001; Station Name: MERIC NEHRI-KAPIKULE	As
3	Station No: 01-11-00-001; Station Name: MERIC NEHRI-KAPIKULE	As
4	Station No: 01-11-00-001; Station Name: MERIC NEHRI-KAPIKULE	As
5	Station No: 01-11-00-001; Station Name: MERIC NEHRI-KAPIKULE	As
6	Station No: 01-11-00-001; Station Name: MERIC NEHRI-KAPIKULE	BOD5
7	Station No: 01-11-00-001; Station Name: MERIC NEHRI-KAPIKULE	BOD5
8	Station No: 01-11-00-001; Station Name: MERIC NEHRI-KAPIKULE	BOD5
9	Station No: 01-11-00-001; Station Name: MERIC NEHRI-KAPIKULE	BOD5
10	Station No: 01-11-00-001; Station Name: MERIC NEHRI-KAPIKULE	BOD5

	Year	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
1	2013-01-01	NA	NA	NA	168.25	NA	156.87	NA	NA	NA	161.21	NA	NA
2	1987-01-01	NA	NA	NA	4.00	NA	1.00	NA	NA	NA	6.00	NA	NA
3	1992-01-01	NA	NA	32.0	NA	NA	NA	NA	NA	NA	NA	NA	NA
4	1994-01-01	NA	NA	NA	NA	NA	9.00	NA	NA	NA	NA	NA	NA
5	2013-01-01	NA	3.49	NA	1.17	NA	52.20	NA	NA	NA	2.74	NA	NA
6	1985-01-01	NA	NA	NA	5.10	NA	6.10	NA	7.2	NA	7.70	NA	7.3
7	1986-01-01	NA	NA	6.5	7.00	NA	4.90	NA	3.6	NA	6.80	NA	9.7
8	1987-01-01	NA	9.20	NA	3.80	NA	4.20	NA	NA	NA	3.80	NA	9.6
9	1988-01-01	NA	9.30	NA	2.70	NA	5.70	NA	7.0	NA	2.10	NA	6.7
10	1989-01-01	NA	5.00	NA	3.60	NA	6.00	NA	6.7	NA	4.50	NA	8.0

Q7 (5 PTS)

- If you look at the data you will see there are missing years in the data.
- For example for station 1, parameter Al is only exists for year 2013 and parameter As only exists for the years 1987, 1992, 1994 and 2013.
- Extend the dataset such that each parameter for each station has row for each year.
- You can do that by using functions `seq.Date()` and `complete()` together.
  - To do this, first generate a vector containing yearly dates from min date to max date with `seq.Date()` function.
  - Next use this vector inside `complete()` function.
  - To better understand how this will work please read the article in the following link. Please Click Here
- Sort the result with `Station Name` and `Parameter`.

Q7 - A Portion of the output

```
> (Q7 %>% as.data.frame())[1:30,2:10]
   Parameter       Year X1 X2 X3     X4 X5     X6 X7
1         Al 1985-01-01 NA NA NA     NA NA     NA NA
2         Al 1986-01-01 NA NA NA     NA NA     NA NA
3         Al 1987-01-01 NA NA NA     NA NA     NA NA
4         Al 1988-01-01 NA NA NA     NA NA     NA NA
5         Al 1989-01-01 NA NA NA     NA NA     NA NA
6         Al 1990-01-01 NA NA NA     NA NA     NA NA
7         Al 1991-01-01 NA NA NA     NA NA     NA NA
8         Al 1992-01-01 NA NA NA     NA NA     NA NA
9         Al 1993-01-01 NA NA NA     NA NA     NA NA
10        Al 1994-01-01 NA NA NA     NA NA     NA NA
11        Al 1995-01-01 NA NA NA     NA NA     NA NA
12        Al 1996-01-01 NA NA NA     NA NA     NA NA
13        Al 1997-01-01 NA NA NA     NA NA     NA NA
14        Al 1998-01-01 NA NA NA     NA NA     NA NA
15        Al 1999-01-01 NA NA NA     NA NA     NA NA
16        Al 2000-01-01 NA NA NA     NA NA     NA NA
17        Al 2001-01-01 NA NA NA     NA NA     NA NA
18        Al 2002-01-01 NA NA NA     NA NA     NA NA
19        Al 2003-01-01 NA NA NA     NA NA     NA NA
20        Al 2004-01-01 NA NA NA     NA NA     NA NA
21        Al 2005-01-01 NA NA NA     NA NA     NA NA
22        Al 2006-01-01 NA NA NA     NA NA     NA NA
23        Al 2007-01-01 NA NA NA     NA NA     NA NA
24        Al 2008-01-01 NA NA NA     NA NA     NA NA
25        Al 2009-01-01 NA NA NA     NA NA     NA NA
26        Al 2010-01-01 NA NA NA     NA NA     NA NA
27        Al 2011-01-01 NA NA NA     NA NA     NA NA
28        Al 2012-01-01 NA NA NA     NA NA     NA NA
29        Al 2013-01-01 NA NA NA 168.25 NA 156.87 NA
30        As 1985-01-01 NA NA NA     NA NA     NA NA
```

```
> (Q7 %>% as.data.frame())[31:60,2:10]
   Parameter       Year X1   X2 X3   X4 X5   X6 X7
31        As 1986-01-01 NA   NA NA   NA NA   NA NA
32        As 1987-01-01 NA   NA NA 4.00 NA  1.0 NA
33        As 1988-01-01 NA   NA NA   NA NA   NA NA
34        As 1989-01-01 NA   NA NA   NA NA   NA NA
35        As 1990-01-01 NA   NA NA   NA NA   NA NA
36        As 1991-01-01 NA   NA NA   NA NA   NA NA
37        As 1992-01-01 NA   NA 32   NA NA   NA NA
38        As 1993-01-01 NA   NA NA   NA NA   NA NA
39        As 1994-01-01 NA   NA NA   NA NA  9.0 NA
40        As 1995-01-01 NA   NA NA   NA NA   NA NA
41        As 1996-01-01 NA   NA NA   NA NA   NA NA
42        As 1997-01-01 NA   NA NA   NA NA   NA NA
43        As 1998-01-01 NA   NA NA   NA NA   NA NA
44        As 1999-01-01 NA   NA NA   NA NA   NA NA
45        As 2000-01-01 NA   NA NA   NA NA   NA NA
46        As 2001-01-01 NA   NA NA   NA NA   NA NA
47        As 2002-01-01 NA   NA NA   NA NA   NA NA
48        As 2003-01-01 NA   NA NA   NA NA   NA NA
49        As 2004-01-01 NA   NA NA   NA NA   NA NA
50        As 2005-01-01 NA   NA NA   NA NA   NA NA
51        As 2006-01-01 NA   NA NA   NA NA   NA NA
52        As 2007-01-01 NA   NA NA   NA NA   NA NA
53        As 2008-01-01 NA   NA NA   NA NA   NA NA
54        As 2009-01-01 NA   NA NA   NA NA   NA NA
55        As 2010-01-01 NA   NA NA   NA NA   NA NA
56        As 2011-01-01 NA   NA NA   NA NA   NA NA
57        As 2012-01-01 NA   NA NA   NA NA   NA NA
58        As 2013-01-01 NA 3.49 NA 1.17 NA 52.2 NA
59         B 1985-01-01 NA   NA NA   NA NA   NA NA
60         B 1986-01-01 NA   NA NA   NA NA   NA NA
```

Q8 (5 PTS)

- Create two columns namely `Station_Code` and `Station_Name` from the column `Station_Name`.

```
> Q8 %>% as.data.frame() %>% head(10)
```

	Station_Code	Station_Name	Parameter	Year	X1	X2	X3	X4	X5	X6	X7
1	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1985-01-01	NA	NA	NA	NA	NA	NA	NA
2	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1986-01-01	NA	NA	NA	NA	NA	NA	NA
3	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1987-01-01	NA	NA	NA	NA	NA	NA	NA
4	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1988-01-01	NA	NA	NA	NA	NA	NA	NA
5	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1989-01-01	NA	NA	NA	NA	NA	NA	NA
6	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1990-01-01	NA	NA	NA	NA	NA	NA	NA
7	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1991-01-01	NA	NA	NA	NA	NA	NA	NA
8	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1992-01-01	NA	NA	NA	NA	NA	NA	NA
9	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1993-01-01	NA	NA	NA	NA	NA	NA	NA
10	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1994-01-01	NA	NA	NA	NA	NA	NA	NA

	X8	X9	X10	X11	X12
1	NA	NA	NA	NA	NA
2	NA	NA	NA	NA	NA
3	NA	NA	NA	NA	NA
4	NA	NA	NA	NA	NA
5	NA	NA	NA	NA	NA
6	NA	NA	NA	NA	NA
7	NA	NA	NA	NA	NA
8	NA	NA	NA	NA	NA
9	NA	NA	NA	NA	NA
10	NA	NA	NA	NA	NA

Q9 (5 PTS)

- Use `pivot_longer` to move `X1:X12` to rows so that months will appear in rows

```
> Q9 %>% as.data.frame() %>% head(10)
```

	Station_Code	Station_Name	Parameter	Year	Month	Value
1	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1985-01-01	X1	NA
2	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1985-01-01	X2	NA
3	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1985-01-01	X3	NA
4	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1985-01-01	X4	NA
5	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1985-01-01	X5	NA
6	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1985-01-01	X6	NA
7	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1985-01-01	X7	NA
8	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1985-01-01	X8	NA
9	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1985-01-01	X9	NA
10	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1985-01-01	X10	NA

Q10 (10 PTS)

- Remove X's in the `Month` column so you get values 1 to 12 in months
- Convert `Month` column to numeric.
- Extract only `Year` part in the `Year` column and assign it to the `Year` column itself.
- Create a new column named `Year_Month` by combining the `Year` and `Month` column.
  - For example, values should be in 1985-02 or 2003-08 format.
  - After combining `Year` and `Month` together, you will notice that, `Year_Month` is in character format.
  - Use `as.yearmon()` in the `zoo` package and `as.Date` together to convert `Year_Month` into the desired Year Month format.

Q10) PART OF OUTPUT

```
> Q10 %>% as.data.frame() %>% head(20)
```

	Station_Code	Station_Name	Parameter	Year	Month	Value	Year_Month
1	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1985	1	NA	1985-01-01
2	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1985	2	NA	1985-02-01
3	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1985	3	NA	1985-03-01
4	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1985	4	NA	1985-04-01
5	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1985	5	NA	1985-05-01
6	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1985	6	NA	1985-06-01
7	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1985	7	NA	1985-07-01
8	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1985	8	NA	1985-08-01
9	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1985	9	NA	1985-09-01
10	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1985	10	NA	1985-10-01
11	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1985	11	NA	1985-11-01
12	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1985	12	NA	1985-12-01
13	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1986	1	NA	1986-01-01
14	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1986	2	NA	1986-02-01
15	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1986	3	NA	1986-03-01
16	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1986	4	NA	1986-04-01
17	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1986	5	NA	1986-05-01
18	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1986	6	NA	1986-06-01
19	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1986	7	NA	1986-07-01
20	01-11-00-001	MERIC NEHRI-KAPIKULE	Al	1986	8	NA	1986-08-01

Q11 (5 PTS)

- Finally, use `pivot_wider` to move each parameter to the columns.
- Sort the output by the `Station_Code`
- The result of `Q11` is the same as the `final_data.xlsx`

```
> Q11 %>% as.data.frame() %>% head(4)
```

	Station_Code	Station_Name	Year	Month	Year_Month	Al	As	B	Ba	BOD5	Ca++
1	01-11-00-001	MERIC NEHRI-KAPIKULE	1985	1	1985-01-01	NA	NA	NA	NA	NA	NA
2	01-11-00-001	MERIC NEHRI-KAPIKULE	1985	2	1985-02-01	NA	NA	NA	NA	NA	NA
3	01-11-00-001	MERIC NEHRI-KAPIKULE	1985	3	1985-03-01	NA	NA	NA	NA	NA	NA
4	01-11-00-001	MERIC NEHRI-KAPIKULE	1985	4	1985-04-01	NA	NA	NA	NA	5.1	76

	Cd	Cl-	CN-	Co	CO2	COD	Col	Cr	Cu	DO	DO%	E-Coli	EC	F-	F-Coli	F-Strp	Fe
1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
3	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
4	NA	24.5	0	NA	NA	NA	NA	0	1	6.4	NA	NA	569	NA	NA	NA	1000

	Fenoller	Hg	Hidrokarbonlar	K+	M-Al	M.Oil	Mg++	Mn	Na+	NH4-N	Ni	NO2-N	NO3-N
1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
3	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
4	NA	NA	NA	6.63	165	35.8	14.6	0	29.9	2	NA	0	2.9

	o-PO4	P-Al	PAH	Pb	pH	pV	Qanlik	Se	SO4=	SS	Surfaktanlar	T	T-Coli	TDS
1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
3	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
4	2.1	NA	NA	NA	8.1	3.28	NA	NA	84	431	NA	11.5	NA	299

	TH	TKN	TOC	Top.N	Top.P	Tot.Pest.	Turb	Zn
1	NA	NA	NA	NA	NA	NA	NA	NA
2	NA	NA	NA	NA	NA	NA	NA	NA
3	NA	NA	NA	NA	NA	NA	NA	NA
4	250	NA	NA	NA	NA	NA	96	NA

Q12 (5 PTS)

- Find min, max, mean and sd values for the parameter BOD5 for each station.

```
> Q12 %>% as.data.frame
```

	Station_Code	Station_Name	mean_BOD5	sd_BOD5	min_BOD5	max_BOD5
1	01-11-00-001	MERIC NEHRI-KAPIKULE	4.229630	2.986572	0	23
2	01-11-00-002	TUNCA NEHRI-SU AKACAGI	4.489933	3.049296	0	16
3	01-11-00-003	ARDA NEHRI-ESKI DEMIRYOLU KOPRUSU	2.238889	2.458612	0	25
4	01-11-00-004	MEKAN DERE SARAY AYVACIK GOLETI AKS YERI	3.800000	3.155243	0	10
5	01-11-00-008	MERIC NEHRI-ESKIKOY	4.271329	2.895860	0	21
6	01-11-00-009	MERIC NEHRI-IPSALA	5.466225	4.125803	0	40
7	01-11-00-011	CORLU SUYU-CERKEZKOY GIRISI	20.821429	23.298678	1	108
8	01-11-00-012	CORLU SUYU-CERKEZKOY CIKISI	90.782667	224.980282	4	1988
9	01-11-00-013	ERGENE NEHRI-CORLU KOPRUSU	7.757143	13.830074	0	96
10	01-11-00-014	ERGENE NEHRI-INANLI	53.127027	41.524838	0	215
11	01-11-00-015	ERGENE NEHRI-LULEBURGAZ	53.233333	45.945184	0	225

Q13 (10 PTS)

- Find min, max, mean and sd values for the parameters `BOD5`, `Ca++`, `Cl-`, `Cu`, `DO`, `EC`, `Fe`, `Mg++` and `Na+` for each station.
- Don't write separate code for each parameter inside the `summarize` function, use the function `across()` instead.

```
> Q13 %>% as.data.frame %>% head(4)
```

	Station_Code	Station_Name	BOD5_mean	BOD5_sd
1	01-11-00-001	MERIC NEHRI-KAPIKULE	4.229630	2.986572
2	01-11-00-002	TUNCA NEHRI-SU AKACAGI	4.489933	3.049296
3	01-11-00-003	ARDA NEHRI-ESKI DEMIRYOLU KOPRUSU	2.238889	2.458612
4	01-11-00-004	MEKAN DERE SARAY AYVACIK GOLETI AKS YERI	3.800000	3.155243

	BOD5_min	BOD5_max	Ca++_mean	Ca++_sd	Ca++_min	Ca++_max	Cl-_mean	Cl-_sd
1	0	23	71.76278	19.58539	1.48	130.00	26.22488	10.89359
2	0	16	78.89114	14.67330	46.00	130.00	52.75383	23.59513
3	0	25	40.46160	10.21230	20.00	68.00	17.25215	10.03040
4	0	10	57.72600	47.92053	10.00	124.34	33.30200	26.74330

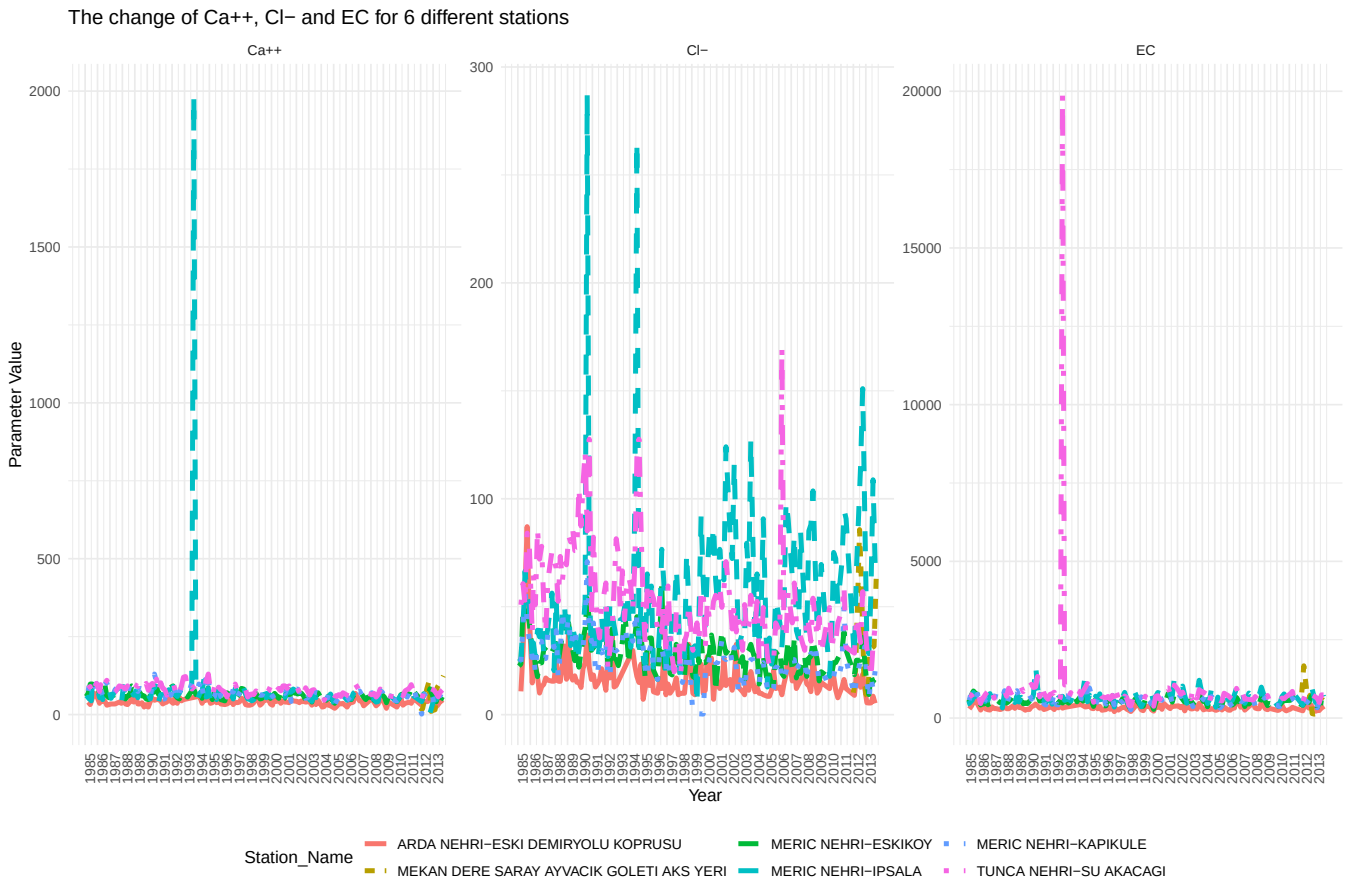
	Cl-_min	Cl-_max	Cu_mean	Cu_sd	Cu_min	Cu_max	DO_mean	DO_sd	DO_min
1	0.26	70.90	270.49950	526.01173	0.00	1870.00	9.699198	2.340753	3.1
2	16.93	170.20	90.75556	160.36788	0.00	600.00	10.391275	2.494668	3.1
3	5.27	87.10	27.20778	65.64731	0.00	200.00	10.108333	2.290006	2.8
4	8.57	85.73	85.68333	71.39003	1.14	182.62	10.870000	1.662027	8.6

	DO_max	EC_mean	EC_sd	EC_min	EC_max	Fe_mean	Fe_sd	Fe_min	Fe_max
1	18.2	609.7531	150.0754	263	1196	967.9350	511.1803	200.0	2000.0
2	16.2	863.6174	1577.9925	376	19885	382.2333	291.9671	0.0	855.9
3	15.4	331.7778	76.1496	205	652	284.0824	239.5492	0.0	700.0
4	13.0	629.1000	465.6192	126	1668	538.0000	364.3594	153.4	1024.9

	Mg++_mean	Mg++_sd	Mg++_min	Mg++_max	Na+_mean	Na+_sd	Na+_min	Na+_max
1	17.451481	6.670011	0.0	48.0	29.52025	12.103190	5.750	79.12
2	27.510811	8.127800	4.9	48.6	41.40097	18.033151	12.199	107.64
3	9.213846	3.987035	2.4	29.2	14.97457	7.356644	2.300	40.25
4	4.790000	3.554793	1.2	11.5	16.60400	11.974782	3.450	44.35

Q14 (10 PTS)

- Create the following graph.



FINAL REMARKS (Due Date: 16/01/2023, 13:00)

- For your take home exam to be graded you have to do two thing.
  1. Copy all your codes to the exam answer sheet provided to you. Don't copy outputs, figures etc. to the exam answer sheet. Then print it out, sign and deliver to the lecturer until the deadline. You should bring the take home exam personally as signature will be taken as a proof.
  2. You should also upload your take home exam (programming script + answer sheet) to YTU online campus system until the deadline.
- If you fail to submit any of the item 1 and 2, your take home exam won't be graded.
- You don't have to use R, but it is recommended. You can also use other programming languages like Python, Matlab, Jupyter Notebook etc.
- The announcement date for the take home exam is **26/12/2022 15:00** and deadline for the homework is **16/01/2023 13:00**.
- You can ask your friends or to the lecturer to get help but don't copy somebody else's code.