

## Tutorial Exercises Week 3

### Question 1

Install the R package "WDI". This is the World Bank Development Indicators R package. It is a way to access the data on <https://data.worldbank.org/> conveniently within R.

Search for the name of indicator which gives the annual inflation in consumer prices (in %). Do this by using the `WDIsearch()` function to search for “inflation”.

Once you have the name of the indicator, use the `WDI()` function to get the annual inflation rate in the Netherlands. Use the following four arguments in the `WDI()` function:

- `indicator` =[the indicator you found in the first step (surrounded by quotes)]
- `country` = "NLD"
- `start` = 1970
- `end` = 2022

What was the inflation rate in the Netherlands in 2022 according to these data?

Write your answer with 3 digits after the decimal

\* Solution

We can install the package with the `install.packages()` function:

```
install.packages("WDI")
```

We then load the package and search for inflation indicators:

```
library(WDI)
WDIsearch("inflation")
```

	indicator	name
8924	FP.CPI.TOTL.ZG	Inflation, consumer prices (annual %)
8926	FP.FPI.TOTL.ZG	Inflation, food prices (annual %)

```

8928      FP.WPI.TOTL.ZG      Inflation, wholesale prices (annual %)
13751    NY.GDP.DEFL.87.ZG      Inflation, GDP deflator (annual %)
13752    NY.GDP.DEFL.KD.ZG      Inflation, GDP deflator (annual %)
13753 NY.GDP.DEFL.KD.ZG.AD Inflation, GDP deflator: linked series (annual %)

```

We can see that the inflation in consumer prices indicator is called "FP.CPI.TOTL.ZG". We then use the `WDI()` function to download the values for the "FP.CPI.TOTL.ZG" indicator for the Netherlands between 1970-2022:

```

df <- WDI(indicator = "FP.CPI.TOTL.ZG", country = "NLD",
          start = 1970, end = 2022)

```

We then subset the data for year 2022 to see the value of inflation in that year:

```

df[df$year == 2022, ]

      country iso2c iso3c year FP.CPI.TOTL.ZG
1 Netherlands   NL   NLD 2022      10.00121

```

## Question 2

What was the average annual inflation rate in consumer prices (in %) over the period 2000-2019?

Write your answer with 3 digits after the decimal.

\* Solution

To do this we can subset the data to years at least as large as 2000 *and* at most 2019:

```

df[df$year >= 2000 & df$year <= 2019, ]

      country iso2c iso3c year FP.CPI.TOTL.ZG
4 Netherlands   NL   NLD 2019      2.6336991
5 Netherlands   NL   NLD 2018      1.7034979
6 Netherlands   NL   NLD 2017      1.3814587
7 Netherlands   NL   NLD 2016      0.3166667
8 Netherlands   NL   NLD 2015      0.6002481
9 Netherlands   NL   NLD 2014      0.9760351
10 Netherlands  NL   NLD 2013      2.5068985
11 Netherlands  NL   NLD 2012      2.4555477
12 Netherlands  NL   NLD 2011      2.3410702
13 Netherlands  NL   NLD 2010      1.2753057
14 Netherlands  NL   NLD 2009      1.1897769
15 Netherlands  NL   NLD 2008      2.4865020

```

16	Netherlands	NL	NLD	2007	1.6138586
17	Netherlands	NL	NLD	2006	1.1015011
18	Netherlands	NL	NLD	2005	1.6881302
19	Netherlands	NL	NLD	2004	1.2636474
20	Netherlands	NL	NLD	2003	2.0919984
21	Netherlands	NL	NLD	2002	3.2875310
22	Netherlands	NL	NLD	2001	4.1558413
23	Netherlands	NL	NLD	2000	2.3605223

To get the average of inflation in those years we can take the mean of the inflation indicator in the subsetting data:

```
mean(df[df$year >= 2000 & df$year <= 2019, ]$FP.CPI.TOTL.ZG)

[1] 1.871487
```

### Question 3

What was the median annual inflation rate in consumer prices (in %) in years that end in a zero between 1970-2020 (so the years 1970, 1980, 1990, 2000, 2010, 2020)?

Write your answer with 3 digits after the decimal.

\* Solution

To get the years that end in a zero we can create a sequence in steps of 10:

```
seq(1970, 2020, by = 10)

[1] 1970 1980 1990 2000 2010 2020
```

We can subset the data using the `%in%` operator with those years. For each year in `df$year`, we check if it matches any of the values in `seq(1970, 2020, by = 10)`:

```
df[df$year %in% seq(1970, 2020, by = 10), ]

  country iso2c iso3c year FP.CPI.TOTL.ZG
3  Netherlands    NL   NLD 2020      1.272460
13 Netherlands    NL   NLD 2010      1.275306
23 Netherlands    NL   NLD 2000      2.360522
33 Netherlands    NL   NLD 1990      2.454092
43 Netherlands    NL   NLD 1980      6.513455
53 Netherlands    NL   NLD 1970      3.668931
```

To get the median inflation in those years, we use the `median()` function:

```
median(df[df$year %in% seq(1970, 2020, by = 10), ]$FP.CPI.TOTL.ZG)
```

```
[1] 2.407307
```

An alternative way we could have done this question is by using the *modulo operator*. The modulo operator gives the remainder when we divide two numbers. For example, if we divide 5 by 2, it's 2 with a remainder of 1:

```
5 %% 2
```

```
[1] 1
```

If we divide 19 by 10, we get 1 with a remainder of 9:

```
19 %% 10
```

```
[1] 9
```

If we divide 20 by 10, we get 2 with a remainder of 0:

```
20 %% 10
```

```
[1] 0
```

The years that end in a zero all have a remainder of 0 when we divide by 10, so we could use this to find the rows of the data with years that end in a zero:

```
df[df$year %% 10 == 0, ]
```

	country	iso2c	iso3c	year	FP.CPI.TOTL.ZG
3	Netherlands	NL	NLD	2020	1.272460
13	Netherlands	NL	NLD	2010	1.275306
23	Netherlands	NL	NLD	2000	2.360522
33	Netherlands	NL	NLD	1990	2.454092
43	Netherlands	NL	NLD	1980	6.513455
53	Netherlands	NL	NLD	1970	3.668931

## Question 4

The `WDI()` function returns a dataframe with the most-recent year first. You want to sort the data so that the years are ascending.

Which of the following commands achieves that goal?

- `df <- df[order(df$year), ]`
- `df <- df[order(df$year, decreasing = TRUE), ]`
- `df <- df[sort(df$year), ]`
- `df <- df[sort(df$year, ascending = TRUE), ]`

Also sort your data by year for the remaining questions.

\* Solution

We use the `order()` function to sort data. The `sort()` sorts the values provided to it, and doesn't provide the indices of the ordered items, which is what we need to sort the data.

We want to sort ascending, which is with `decreasing = FALSE`. This is the default in the `order()` function, so we can use the following to sort the data:

```
df <- df[order(df$year), ]
```

## Question 5

Create a vector called `inflation_change` which is the change in inflation from the previous year:

- The first element should be the difference in the inflation rate between 1971 and 1970 (inflation 1971 minus inflation 1970).
- The second element should be the difference in the inflation rate between 1972 and 1971.
- ...
- The last element should be difference in the inflation rate between 2022 and 2021.

This vector should have 52 elements. We don't have a change for 1970 because we don't observe the inflation rate in 1969.

What is the value of the 5th element in your vector? Report your answer with 3 numbers after the decimal.

\* Solution

The first element should be the difference between 1971 and 1970. This is:

```
df$FP.CPI.TOTL.ZG[2] - df$FP.CPI.TOTL.ZG[1]
```

```
[1] 3.808748
```

because 1970 is the 1st element and 1971 is the 2nd.

The second element should be the difference between 1972 and 1971. This is:

```
df$FP.CPI.TOTL.ZG[3] - df$FP.CPI.TOTL.ZG[2]
```

```
[1] 0.3251135
```

because 1971 is the 2nd element and 1972 is the 3rd. The last element should be the difference between 2022 and 2021. This is:

```
df$FP.CPI.TOTL.ZG[53] - df$FP.CPI.TOTL.ZG[52]
```

```
[1] 7.325488
```

because 2021 is the 62nd element and 2022 is the 63rd.

Seeing the pattern here, we can get all the elements together using:

```
inflation_change <- df$FP.CPI.TOTL.ZG[2:53] - df$FP.CPI.TOTL.ZG[1:52]
```

We can check that this has 52 elements:

```
length(inflation_change)
```

```
[1] 52
```

We need to provide the 5th element to answer the question. This is:

```
inflation_change[5]
```

```
[1] 0.6260531
```

## Question 6

You want to add this `inflation_change` vector as a variable to your dataframe which is sorted ascending by year.

However, because your vector has only 52 elements (because we do not know the value of inflation in 1969), we cannot add it directly. Instead, we replace the value for the change of inflation in 1970 with a missing value indicator `NA`.

Which of the following commands will add the `inflation_change` variable correctly in your data?

- `df$inflation_change <- c(inflation_change, NA)`
- `df$inflation_change <- c(NA, inflation_change)`
- `df$inflation_change <- inflation_change`

\* Solution

We don't know what the inflation change is for 1970 because we don't know what inflation was in 1969. Therefore we need to set the value for this year to missing if we add this vector to our dataframe. We add an `NA` to the start (the value in 1970) and then include the change for all the other years:

```
df$inflation_change <- c(NA, inflation_change)
```