

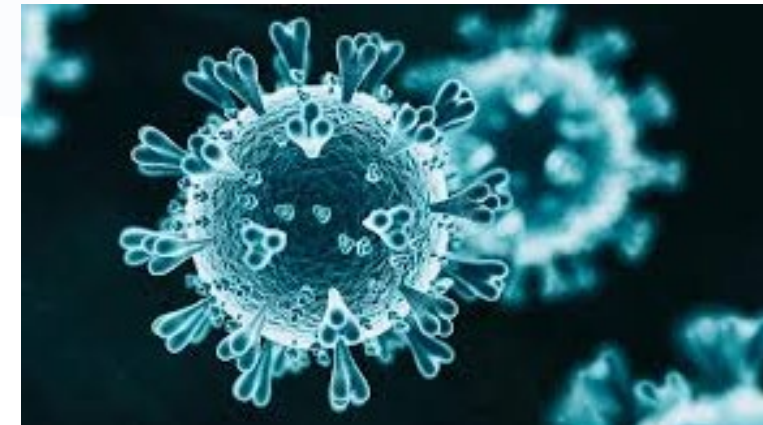
Media & Data Science

Lecturer: Changjun Lee
changjunlee@hanyang.ac.kr

College of Communication
Media Informatics



Practice with the real-world data



[Add your blog!](#)

[Learn R](#)

[R jobs](#) ▾

[Contact us](#)

Tidying the new Johns Hopkins Covid-19 time-series datasets

March 23, 2020

By An Accounting and Data Science Nerd's Corner

 Like 461

[Share](#)

 Tweet

 Share

[This article was first published on [An Accounting and Data Science Nerd's Corner](#), and kindly contributed to R-bloggers]. (You can report issue about the content on this page [here](#))

Want to share your content on R-bloggers? [click here](#) if you have a blog, or [here](#) if you don't.

 Share

 Tweet

SEARCH R-BLOGGERS

[Go](#)

MOST VISITED ARTICLES OF THE WEEK

1. [Google Big Query with R](#)
2. [5 Ways to Subset a Data Frame in R](#)
3. [Covid 19 Tracking](#)
4. [How to write the first for loop in R](#)
5. [Tidying the John Hopkins Covid-19 datasets](#)
6. [Simulating COVID-19 interventions with R](#)
7. [Survey Results: What Degree is Best for Data Science?](#)
8. [How to create a simple Corona dashboard specific to your country](#)
9. [COVID-19 Tracker: Days since](#)

↳ my old blog post about tidying Johns Hopkins CSSE

Download the data for class

Week 2: R Basic Syntax (1)

- Date: 20230309
- Pre-class: Basic syntax, Vector, Array
- Class: Hands-on practice [PDF](#)
 - Data in use: [COV19 data](#)

Matrix 1: Confirmed cases

	2020-02-01	...	2020-10-01	...	2022-09-01
China	11891	...	90567	...	2510703
Italy					
Japan					
Korea					
Spain					
UK					
US	8	...	7279272	...	94665567

Matrix 2: Death cases

	2020-02-01	...	2020-10-01	...	2022-09-01
China	18	
Italy					
Japan					
Korea					
Spain					
UK					
US	0	

7 rows and 32 columns

Access to the matrix

1. UK's total confirmed cases on 2022-09-01
2. South Korea's total confirmed cases on 2022-09-01
3. China's total confirmed cases on 2022-09-01
4. South Korea's increasing confirmed cases on 2022-09-01 compared to the previous month
5. Japan's increasing confirmed cases on 2022-09-01 compared to the previous month

Vector 1: GDP

```
country.name<-c("China","Italy","Japan","Korea","Spain","UK","US")
GDP<-c(12237700479375,
1943835376342,
4872415104315,
1530750923149,
1314314164402,
2637866340434,
19485394000000)
names(GDP)<-country.name
```

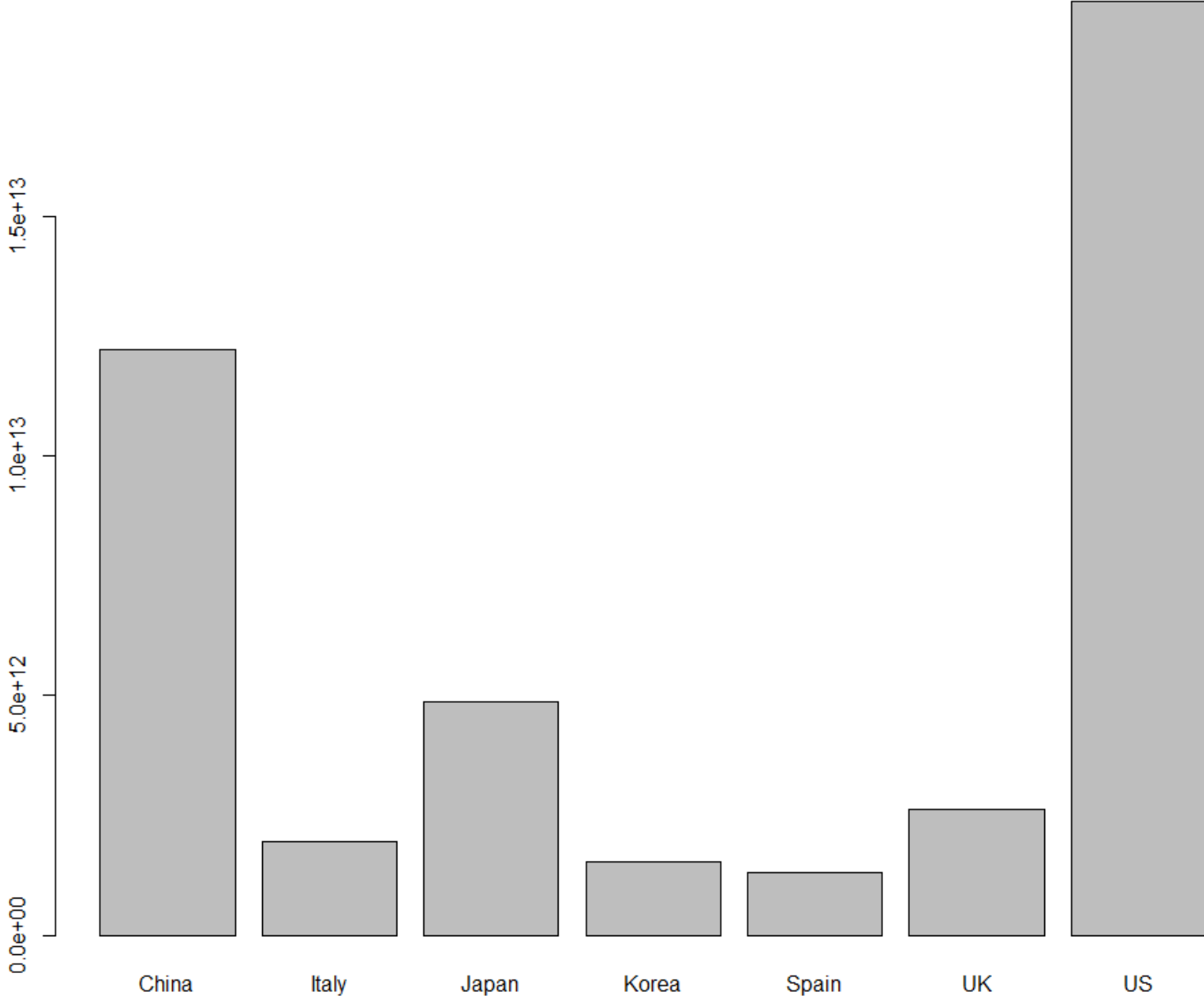
Vector 3: Population Density

```
country.name<-c("China","Italy","Japan","Korea","Spain","UK","US")
pop.density<-c(148, 205, 347, 530, 94, 275, 36)
names(pop.density)<-country.name
```

Vector 2: Population

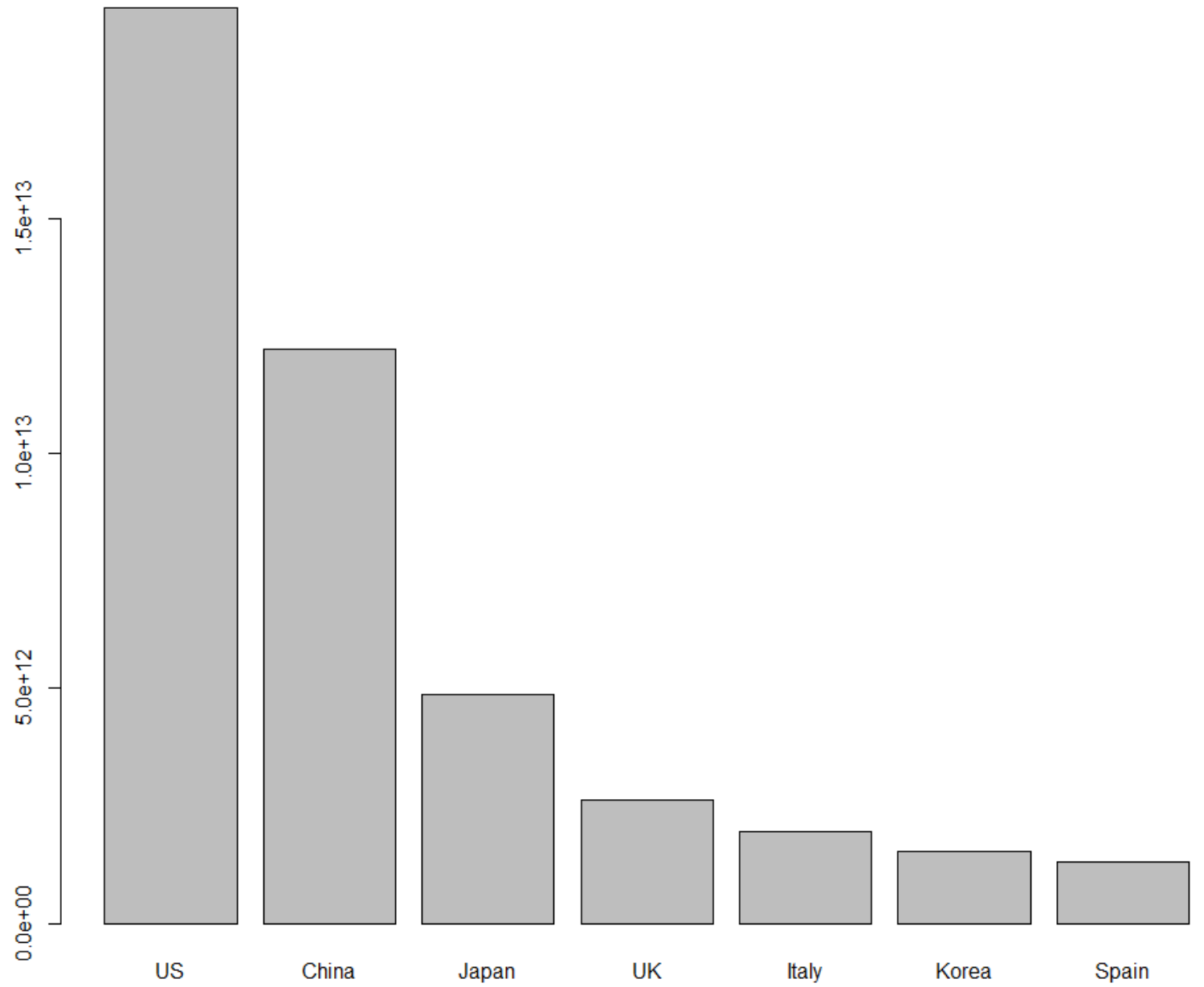
```
pop<-c(1439323776, 60461826, 126476461, 51269185, 46754778,
67886011, 331002651)
names(pop)<-country.name
```

barplot(GDP)




```
barplot(sort(GDP))
```

```
barplot(sort(GDP, decreasing = T))
```



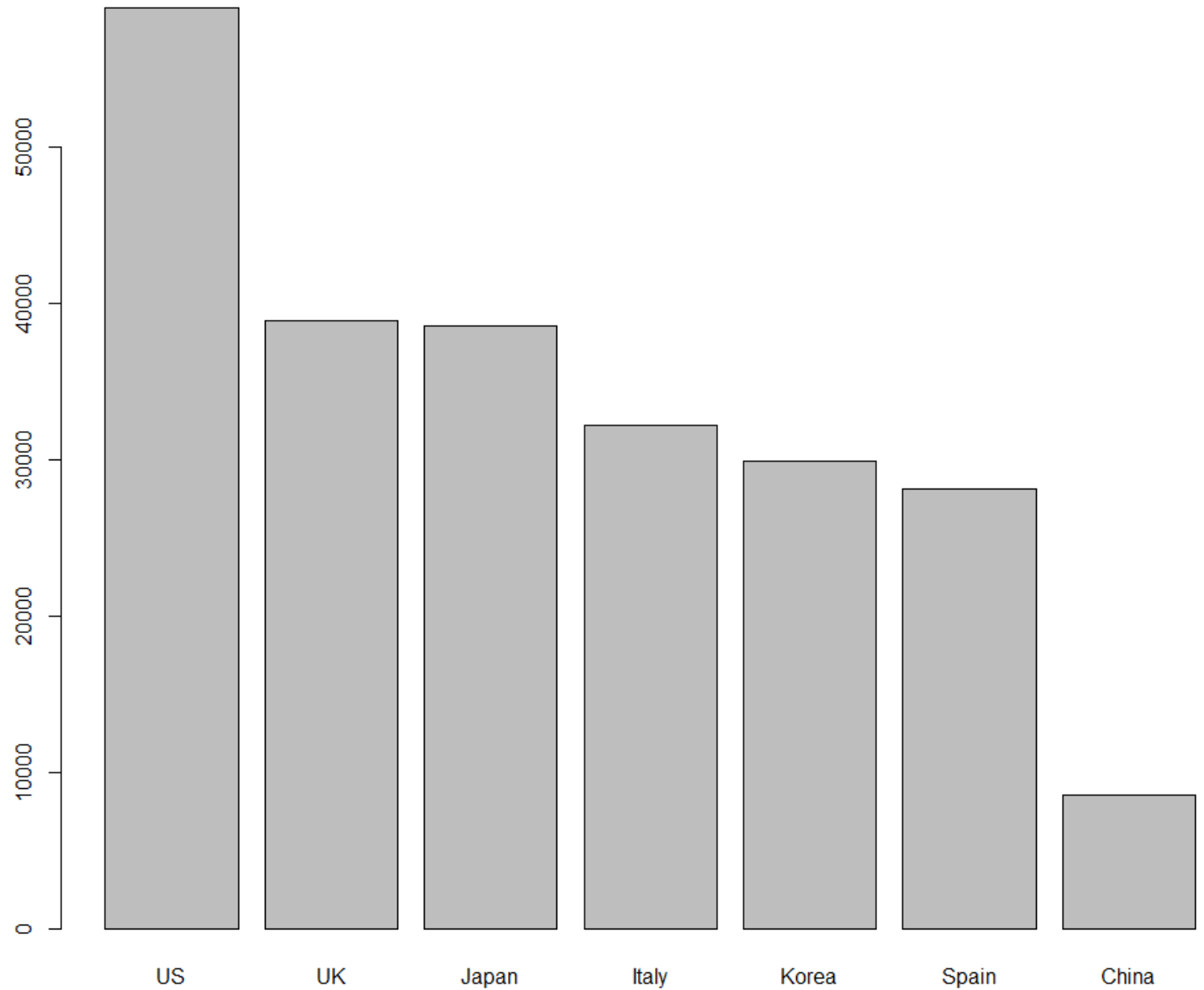
Bar graph of GDP per capita

We have

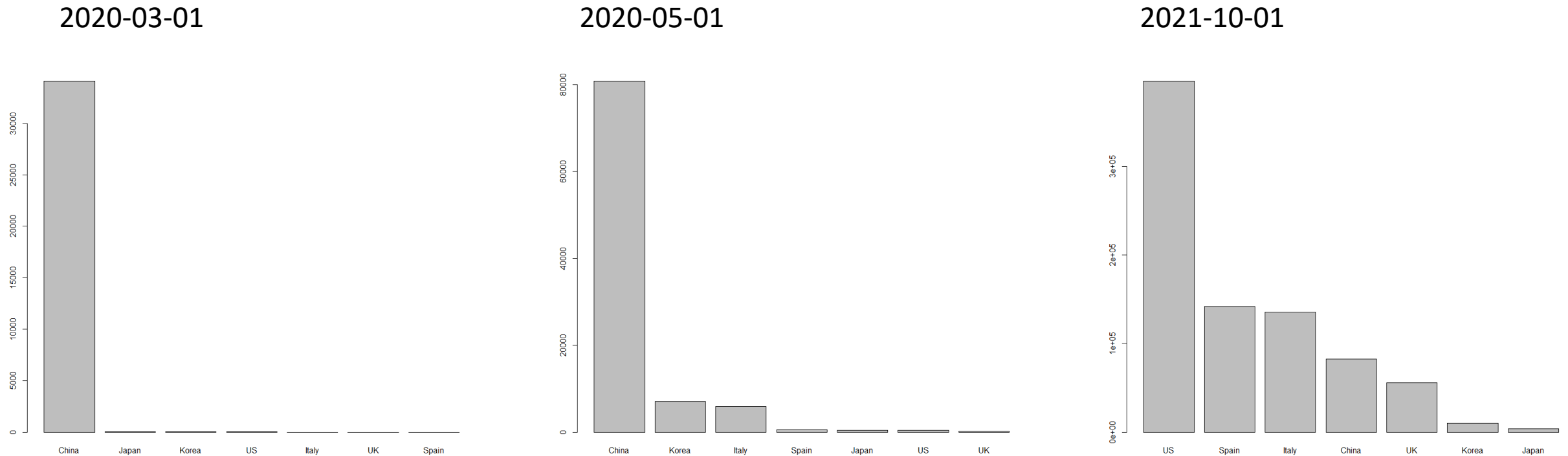
- 1. GDP vector
- 2. Population vector

We know

- 1. `barplot()`
- 2. Vector calculation
- 3. `Sort()`
- 4. `Decreasing=T` option



Confirmed cases bar graph



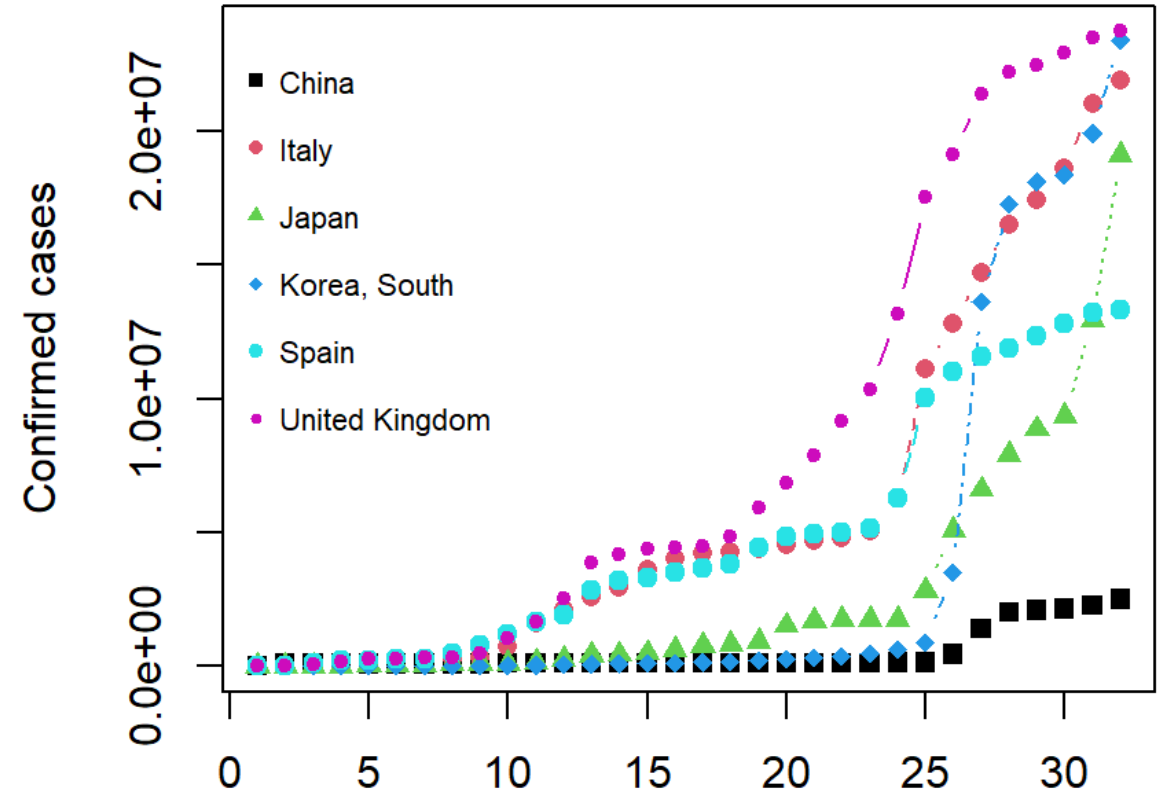
See which columns indicate the dates above (2020-03-01 / 2020-05-01 / 2021-10-01) in the matrix

- `Mat.name[,]`

Use `barplot()` , `sort()` , `decreasing=T` option

matplotlib

	2020-01-23	...	2020-02-15	...	2020-04-07
China	643	...	64813	...	82718
Italy					
Japan					
Korea					
Spain					
UK					
US	1	...	13	...	396223



matplot

First, try

```
matplot(m.conf.case)
```

You must transpose the matrix

```
t(m.conf.case)
```

	2020-01-23	...	2020-02-15	...	2020-04-07
China	643	...	64813	...	82718
Italy					
Japan					
Korea					
Spain					
UK					
US	1	...	13	...	396223

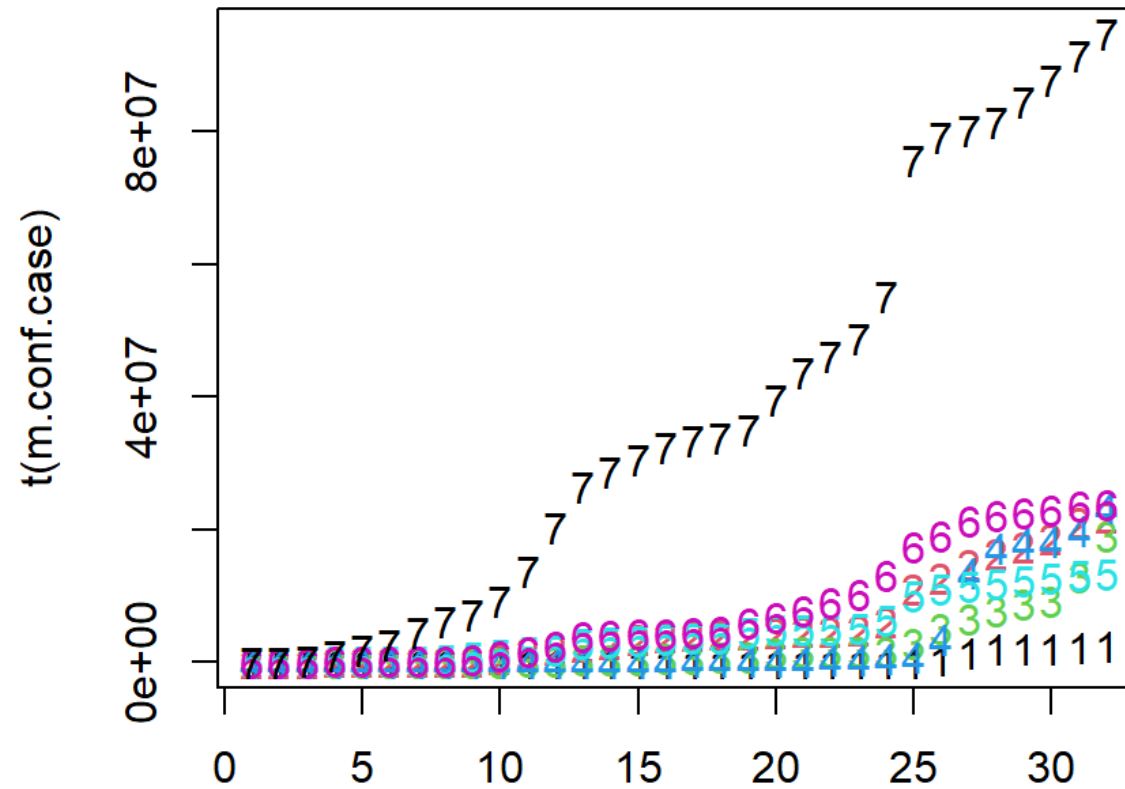


	China	...	Korea	...	US
2020-01-23	643	...	64813	...	82718
...					
...					
...					
...					
...					
2020-04-07	1	...	13	...	396223

matplot

Let's try

```
matplot(t(m.conf.case))
```



matplot

This is the cascade

```
matplot(t(m.conf.case))
```

```
matplot(t(m.conf.case), type='b')
```

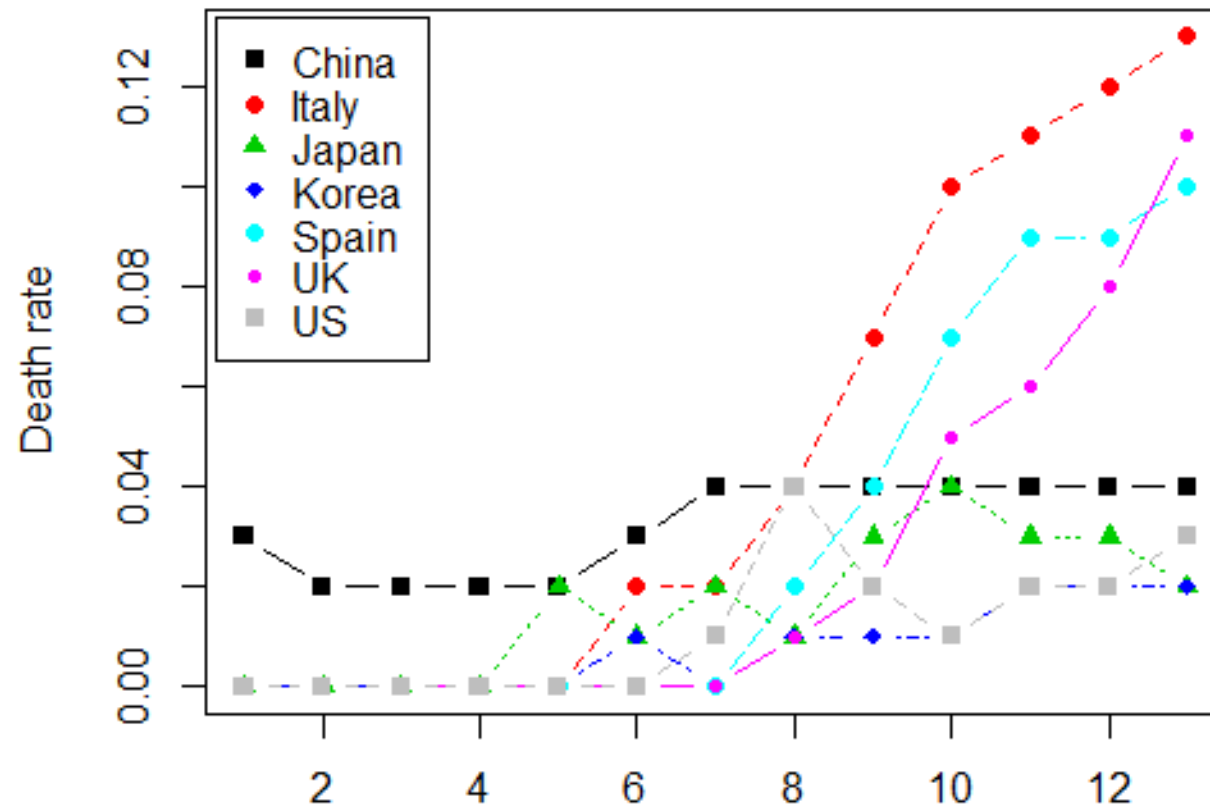
```
matplot(t(m.conf.case), type='b', pch=15:20)
```

```
matplot(t(m.conf.case), type='b', pch=15:20, col=c(1:6, 8),  
ylab="Confirmed cases")
```

```
legend("topleft", inset=0.01, legend=country.name, pch=15:20,  
col=c(1:6, 8), horiz=F)
```

matplot

Try the same graph but now use the death rate

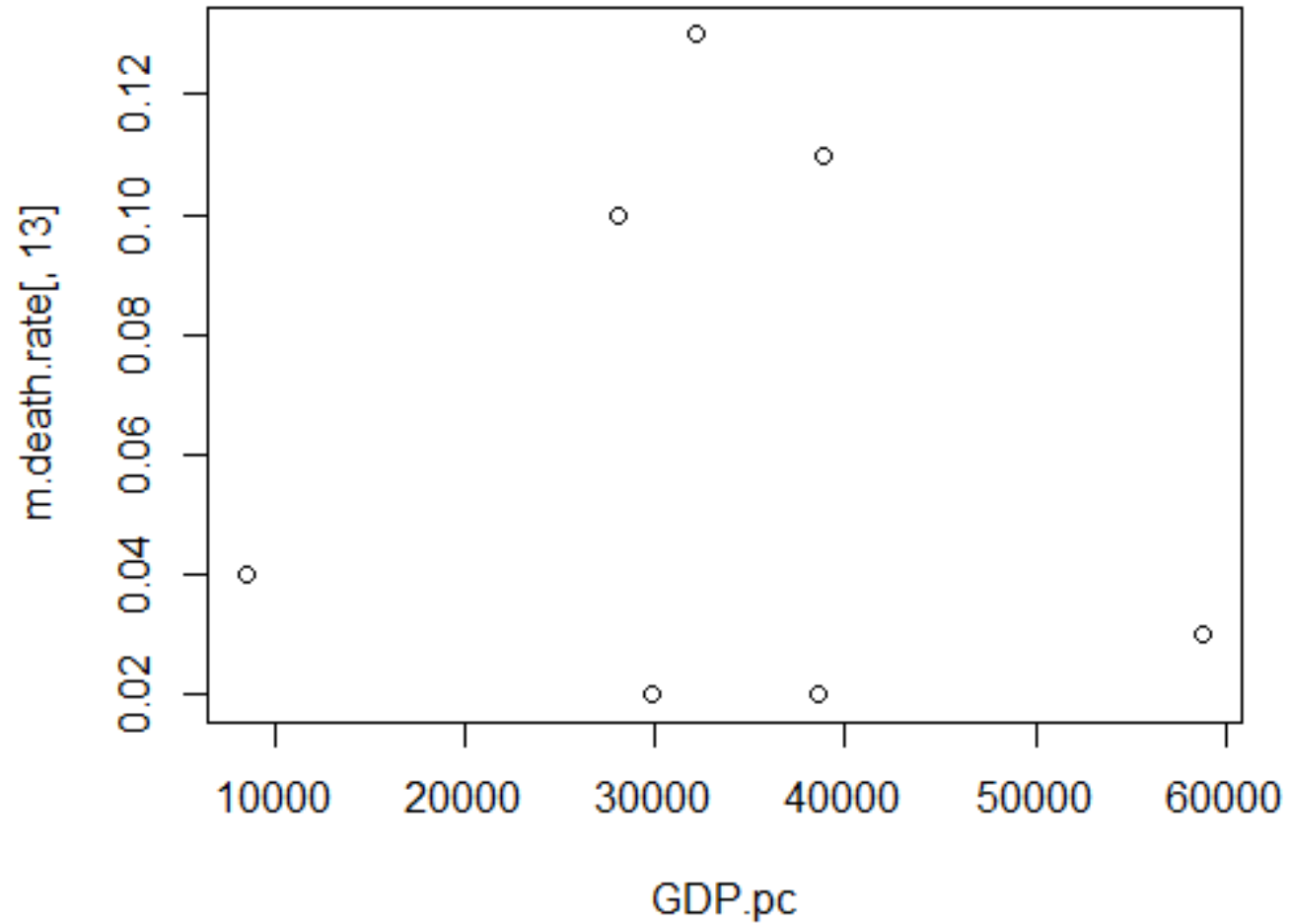


County's wealth and COVID19

I'm now curious about the relationship between countries' GDP per capita and the death rate at the latest time

plot

```
plot(GDP.pc, m.death.rate[,16])
```




plot

```
plot(GDP.pc, m.death.rate[,32])
```


```
plot(GDP.pc, m.death.rate[,32], ylab="Death rate")
```

```
text(GDP.pc, m.death.rate[,32], row.names(m.death.rate),  
cex=1, pos=4, col="blue")
```


Let's change
this to 2,3,4

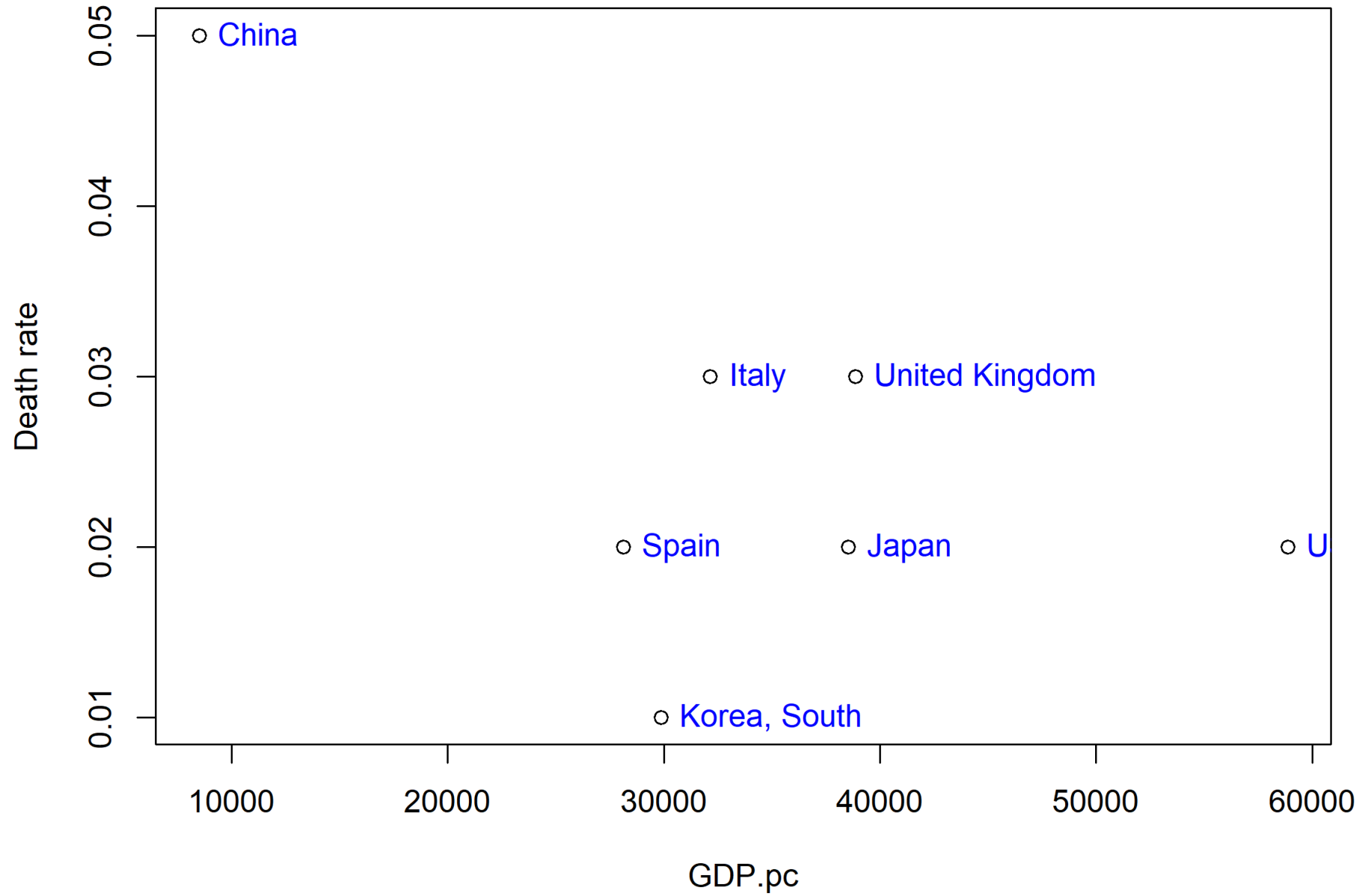


Let's change
this to 1,2,3



Let's change
this to "red",
"black"





County's pop density and COVID19

Let's explore the relationship between countries' population density and the death rate at the latest time

Increasing rate!

When you wonder if the number of confirmed cases is in increasing phase or flattened..

Increasing rate!

	2020-02-01	...	2020-10-01	...	2022-09-01
China					
Italy					
Japan					
Korea					
Spain					
UK					
US					

`m1<-m.conf.case[,-1]`

	2020-02-01	...	2020-10-01	...	2022-09-01
China					
Italy					
Japan					
Korea					
Spain					
UK					
US					

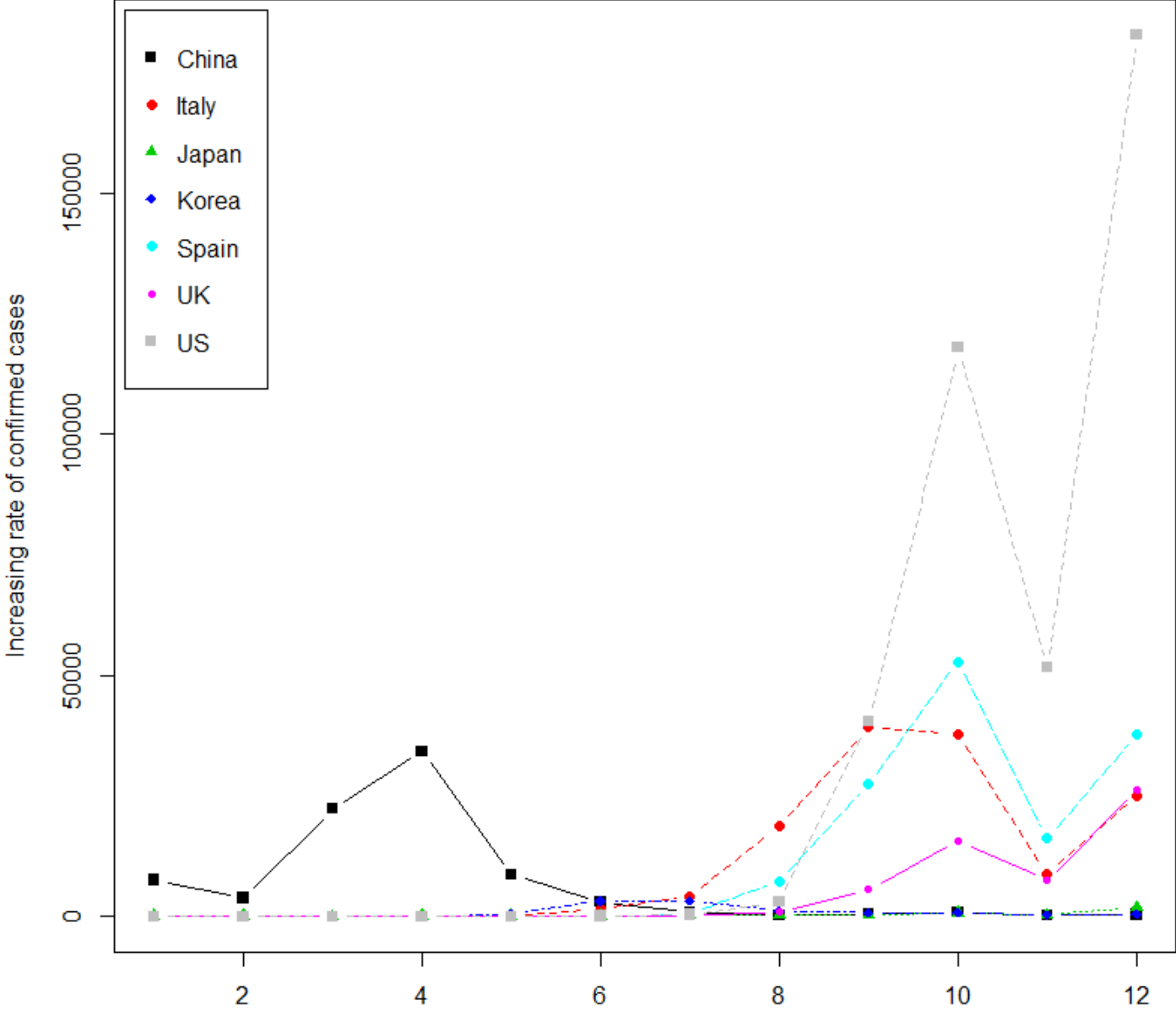
`m2<-m.conf.case[,-32]`

`m3<-m1-m2`

Increasing rate of confirmed cases

Let's visualize like an example →

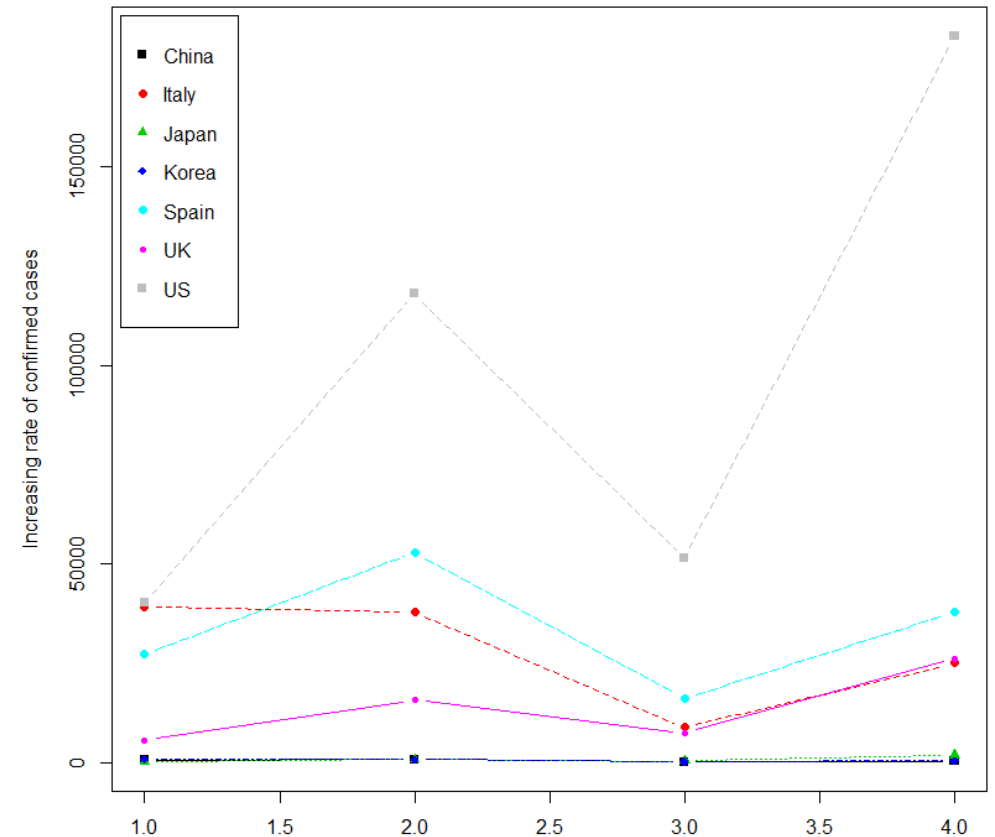
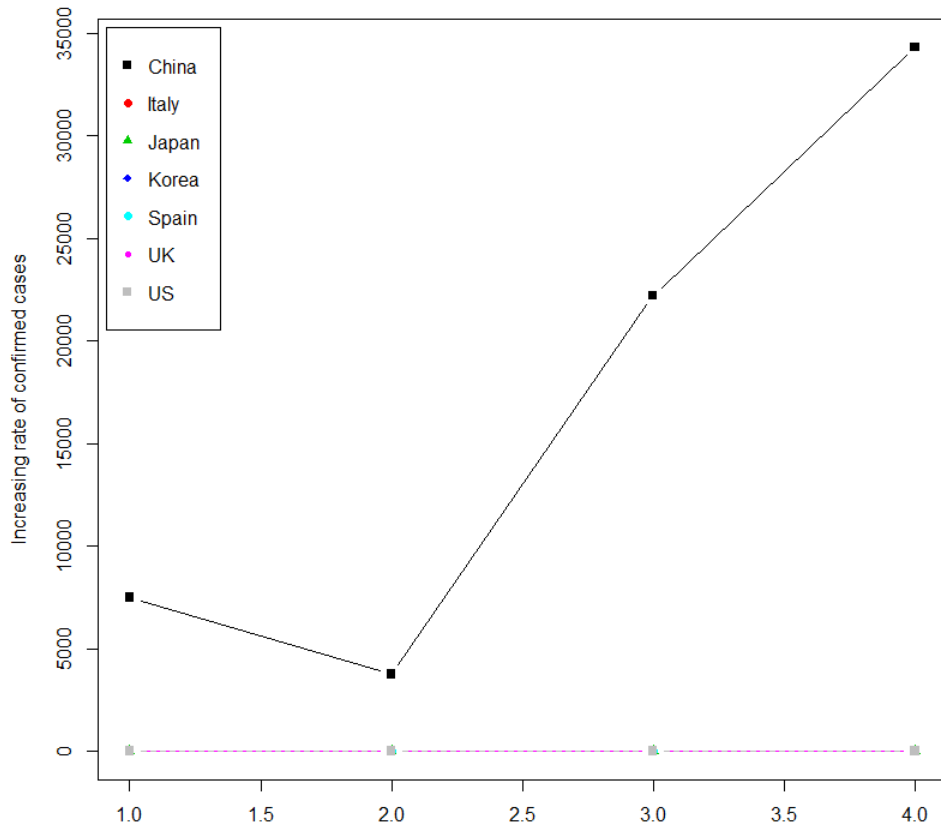
5 mins



5 mins

Increasing rate of confirmed cases

Let's also visualize the first four periods and the last (recent) four periods

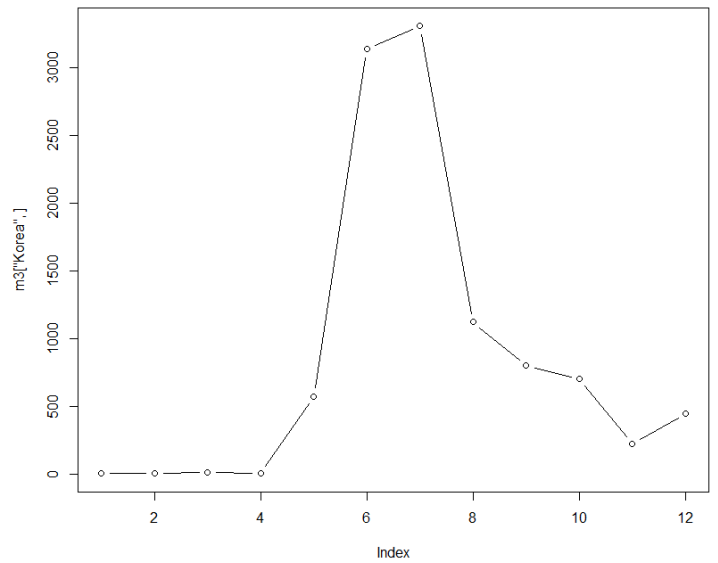


10 mins

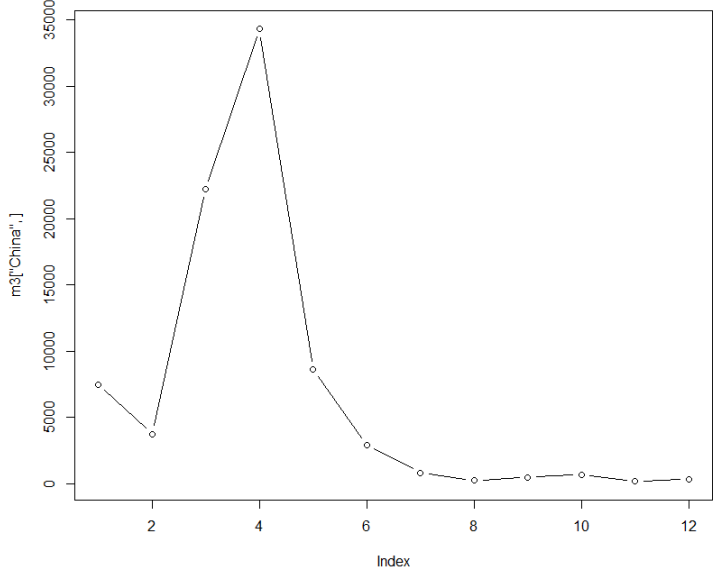
Increasing rate of confirmed cases

Can you also do for the specific countries like Korea, China, and Japan?

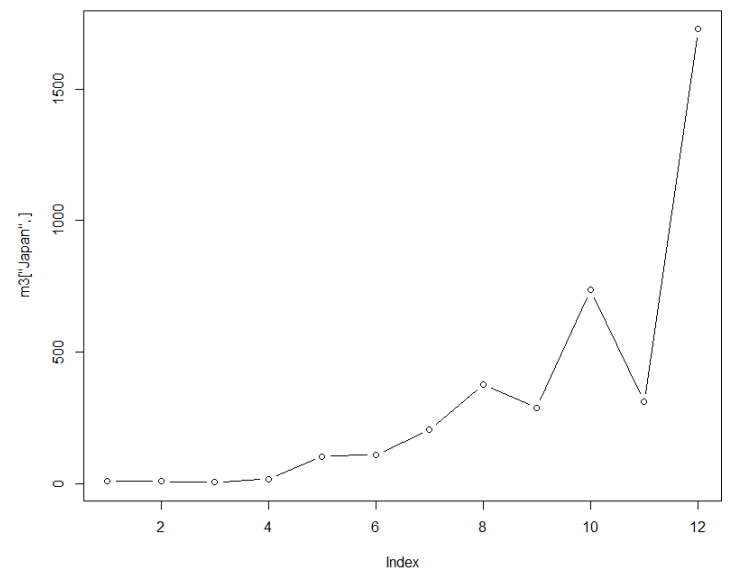
Korea



China

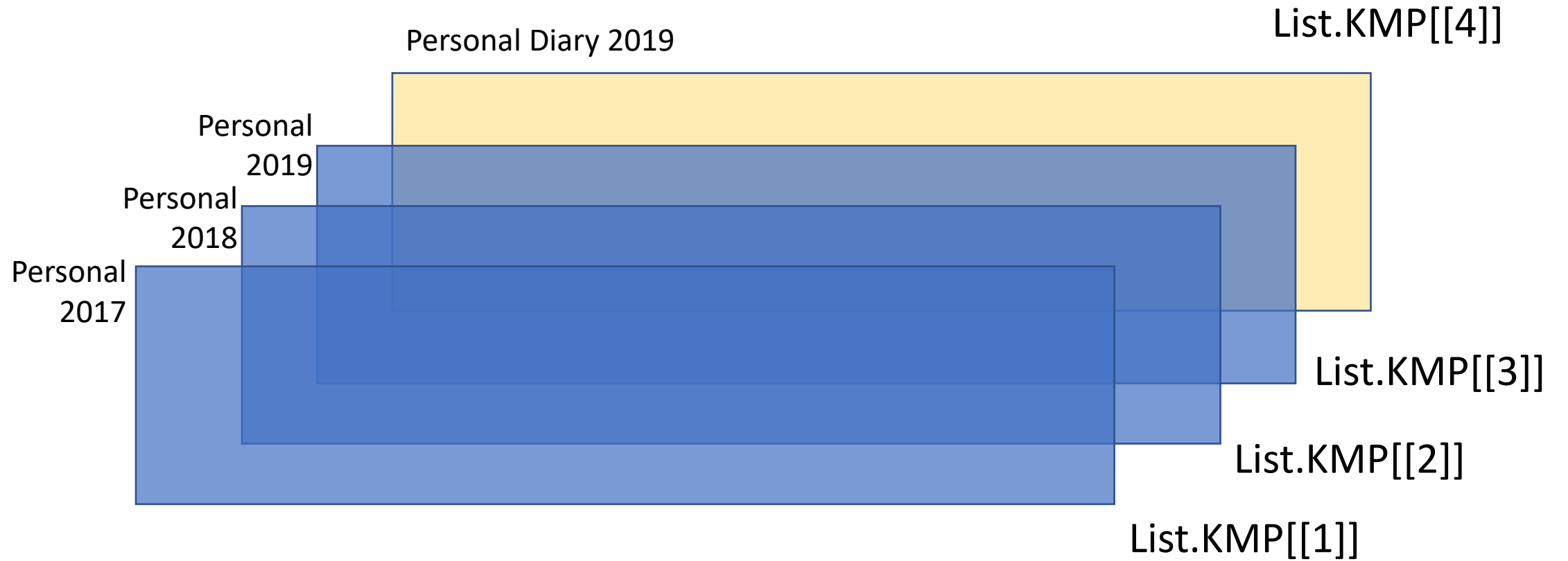


Japan



Practice with Media Panel Data

List.KMP



List.KMP

Personal Diary 2019

Personal
2019

	pid	Smarphone.time	SNS.time	year
1	9920004	0	0	2019
2	12500003	60	0	2019
3	17350004	345	10	2019
4	17670001	90	0	2019
5	23860001	40	0	2019
6	24450001	65	0	2019

Personal
2018

	pid	age	gender	Mobile.tv	Telecom	Smartphone.brand	sp.mobile	sp.device	sp.online.content	sp.offline.contet	year
1	9920004	11	Male	No Smartphone	No Smartphone	No Smartphone	0	0	0	0	21 2019
2	12500003	28	Male	LTE	LG U+	Apple	60	30	90	0	0 2019
3	17350004	14	Female	LTE-A	LG U+	Samsung	21	0	21	0	0 2019
4	17670001	57	Male	LTE	SKT	Samsung	30	0	30	0	20 2019
5	23860001	71	Male	No Smartphone	No Smartphone	No Smartphone	0	0	0	0	0 2019

Personal
2017

	pid	age	gender	Mobile.tv	Telecom	Smartphone.brand	sp.mobile	sp.device	sp.online.content	sp.offline.contet	year
1	9920004	10	Male	No Smartphone	No Smartphone	No Smartphone	0	0	0	8	2018
2	12500003	27	Male	LTE-A	LG U+	Samsung	90	60	359	120	2018
3	17350004	13	Female	LTE-A	LG U+	LG	20	35	55	0	2018
4	17670001	56	Male	LTE	KT	Samsung	39	0	39	0	2018
5	23860001	71	Male	No Smartphone	No Smartphone	No Smartphone	30	0	30	0	2018
6	24450001	58	Male	LTE	SKT	Samsung	65	0	50	60	2018
7	27570001	50	Male	LTE	SKT	Samsung	42	0	53	22	2017
8	17350004	12	Female	LTE	LG U+	Samsung	19	0	19	0	2017
9	17670001	55	Male	LTE	SKT	Samsung	38	0	38	0	2017
10	23860001	70	Male	No Smartphone	No Smartphone	No Smartphone	18	0	18	100	2017
11	24450001	58	Male	LTE	SKT	Samsung	65	0	50	60	2017
12	27570001	50	Male	LTE	SKT	Samsung	32	0	32	36	2017
13	53620001	68	Male	LTE	SKT	Samsung	58	10	68	0	2017
14	59570001	39	Male	LTE	KT	Samsung	63	0	63	25	2017
15	65840001	37	Female	LTE-A	KT	LG	54	10	114	40	2017

List.KMP

Personal 2017

	pid	age	gender	Mobile.tv	Telecom	Smartphone.brand	sp.mobile	sp.device	sp.online.content	sp.offline.contet	year
1	9920004	9	Male	No Smartphone	No Smartphone	No Smartphone	0	0	0	10	2017
2	12500003	26	Male	LTE	KT	Samsung	42	0	53	22	2017
3	17350004	12	Female	LTE	LG U+	Samsung	19	0	19	0	2017
4	17670001	55	Male	LTE	SKT	Samsung	38	0	38	0	2017
5	23860001	70	Male	No Smartphone	No Smartphone	No Smartphone	18	0	18	0	2017
6	24450001	58	Male	LTE	SKT	Samsung	65	0	65	50	2017
7	27570001	50	Male	LTE	SKT	Samsung	32	0	32	36	2017
8	53620001	68	Male	LTE	SKT	Samsung	58	10	68	0	2017
9	59570001	39	Male	LTE	KT	Samsung	63	0	63	25	2017
10	65840001	37	Female	LTE-A	KT	LG	54	10	114	40	2017

List.KMP

Diary 2019

	pid	Smarphone.time	SNS.time	year
1	9920004	0	0	2019
2	12500003	60	0	2019
3	17350004	345	10	2019
4	17670001	90	0	2019
5	23860001	40	0	2019
6	24450001	65	0	2019
7	27570001	170	0	2019
8	53620001	95	0	2019
9	59570001	85	0	2019
10	65840001	75	0	2019

Load List.KMP

The screenshot displays the RStudio interface with the following components:

- Source Editor:** Contains R code for data processing, including filtering by year, summarizing data by group, and saving the resulting list to a file named "List_KMP.RData".
- Environment:** Shows a list of objects created in the global environment, such as "List.KMP", "m.conf.case", "m.death.case", "m1", "m2", "m3", "matched_etry_names", "p17", "p17.df", "p17.s", "p1789", "p18", "p18.s", and "p19".
- Files:** A file explorer window is open, showing the directory structure of the project. A "More" button is circled in white.
- Console:** Displays the output of the R code, showing the structure of the data frames and the successful execution of the save command.

```
152 final.sample.10 %>% filter(year==2018) -> List.KMP[[2]]
153 final.sample.10 %>% filter(year==2019) -> List.KMP[[3]]
154
155
156 # smartphone using time
157 d19$d19M_time19
158 # sns using time
159 d19$d19A_time21
160
161 d19 %>% select(pid, d19M_time19, d19A_time21) %>%
162   group_by(pid) %>% summarize_all(mean, na.rm=T) %>%
163   setNames(c("pid", "Smartphone.time", "SNS.time")) %>%
164   mutate(year=2019) %>% filter(pid %in% chosen.pid) -> d19.s
165 as.data.frame(d19.s) -> List.KMP[[4]]
166
167
168 List.KMP[[4]]
169
170 save(List.KMP, file="List_KMP.RData")
171
172
173 str(List.KMP)
174
175 names(List.KMP)
176 names(List.KMP)<-c("p17", "p18", "p19", "d19")
177 p17.df<-List.KMP[["p17"]]
```

```
..$ sp.mobile      : num [1:10] 0 90 20 39 30 80 33 36 40 59
..$ sp.device      : num [1:10] 0 60 35 0 0 15 0 10 10 12
..$ sp.online.content: num [1:10] 0 359 55 39 30 95 33 46 50 71
..$ sp.offline.contet: num [1:10] 8 120 0 0 0 12 20 0 0 100
..$ year           : num [1:10] 2018 2018 2018 2018 2018 ...
$: 'data.frame':   10 obs. of  11 variables:
..$ pid           : num [1:10] 9920004 12500003 17350004 17670001 23860001 ...
..$ age           : num [1:10] 11 28 14 57 72 60 52 70 41 39
..$ gender        : Factor w/ 2 levels "Male","Female": 1 1 2 1 1 1 1 1 1 2
..$ Mobile.lv     : Factor w/ 5 levels "3G","LTE","LTE-A",...: 5 2 3 2 2 2 2 2 3 3
..$ Telecom       : Factor w/ 5 levels "SKT","KT","LG U+",...: 5 3 3 1 2 1 1 2 3 3
..$ Smartphone.brand : Factor w/ 8 levels "Samsung","Apple",...: 8 2 1 1 1 1 1 3 3 3
..$ sp.mobile      : num [1:10] 0 60 21 30 32 78 35 54 73 45
..$ sp.device      : num [1:10] 0 30 0 0 0 0 0 20 15
..$ sp.online.content: num [1:10] 0 90 21 30 32 78 35 54 93 60
..$ sp.offline.contet: num [1:10] 21 0 0 20 0 24 90 0 20 60
..$ year           : num [1:10] 2019 2019 2019 2019 2019 ...
$: 'data.frame':   10 obs. of  4 variables:
..$ pid           : num [1:10] 9920004 12500003 17350004 17670001 23860001 ...
..$ Smartphone.time: num [1:10] 0 60 345 90 40 65 170 95 85 75
..$ SNS.time       : num [1:10] 0 0 10 0 0 0 0 0 0 0
..$ year           : num [1:10] 2019 2019 2019 2019 2019 ...
> save(List.KMP, file="List_KMP.RData")
>
```

Name	Size	Modified
..		
.RData	339.5 KB	Apr 9, 2020, 12:40 AM
.Rhistory	23 KB	Apr 15, 2020, 6:07 AM
.Rproj.user		
CJ_course_foundationsDS.Rproj	218 B	Apr 15, 2020, 9:11 PM
Code for lecture.R	37.7 KB	Mar 25, 2020, 12:14 PM
cov19.R	18.7 KB	Apr 9, 2020, 9:28 PM
jh_covid19_data_2020-04-07.csv	398.4 KB	Apr 7, 2020, 9:03 PM
kmp		
Lecture code in R(E).R	37.2 KB	Apr 15, 2020, 9:11 PM
Lecture code.R	37.7 KB	Mar 24, 2020, 1:50 AM
Mat_conf_case.RData	572 B	Apr 8, 2020, 8:32 PM
Mat_death_case.RData	465 B	Apr 8, 2020, 8:32 PM
students.csv	114 B	Apr 5, 2020, 5:52 PM
students1.txt	112 B	Apr 5, 2020, 5:51 PM
students2.csv	116 B	Apr 5, 2020, 5:52 PM
students2.txt	114 B	Apr 5, 2020, 5:52 PM
students3.txt	130 B	Apr 5, 2020, 5:53 PM

Load List.KMP

The screenshot displays the RStudio interface with several components:

- Source Editor:** Contains R code for loading and processing data. The code includes filtering by year (2018 and 2019), selecting specific variables, summarizing data by group, and saving the results into a list object named `List.KMP`.
- Environment Pane:** Shows the `List.KMP` object as a list of 4 data frames. The first three elements are `m.conf.case`, `ath.case`, and `ath.rate`, each with 10 observations and 4 variables. The fourth element is `hed_ctry_names`, with 184 observations and 4 variables.
- Terminal:** Shows the output of the R code, displaying the structure of the data frames and the variables within them.
- File Explorer:** A window titled "CJ_course_foundationsDS" is open, showing a file tree. The file `7.csv` is highlighted, and a context menu is open over it. The menu options are: Copy..., Copy To..., Move..., Set As Working Directory, Go To Working Directory, Show Folder in New Window (circled in white), and Show Hidden Files.

Load List.KMP

The screenshot shows the RStudio interface. A file explorer window is open, displaying the contents of the folder 'CJ_course_foundationsDS'. The console window shows the execution of R code, including the command `save(List.KMP, file="List_KMP.Rdata")`.

File Explorer Contents:

이름	수정한 날짜	유형	크기
.Rproj.user	2020-02-18 오전 10:12	파일 폴더	
kmp	2020-04-16 오전 12:21	파일 폴더	
.RData	2020-04-09 오전 12:40	R Workspace	340KB
.Rhistory	2020-04-15 오전 6:07	RHISTORY 파일	24KB
CJ_course_foundationsDS	2020-04-15 오후 9:11	R Project	1KB
Code for lecture	2020-03-25 오후 12:14	R 파일	38KB
cov19	2020-04-09 오후 9:28	R 파일	19KB
jh_covid19_data_2020-04-07	2020-04-07 오후 9:03	Microsoft Excel Com...	399KB
Lecture code in R(E)	2020-04-15 오후 9:11	R 파일	38KB
Lecture code	2020-03-24 오전 1:50	R 파일	38KB
Mat_conf_case	2020-04-08 오후 8:32	R Workspace	1KB
Mat_death_case	2020-04-08 오후 8:32	R Workspace	1KB
students	2020-04-05 오후 5:52	Microsoft Excel Com...	1KB
students1	2020-04-05 오후 5:51	텍스트 문서	1KB
students2	2020-04-05 오후 5:52	Microsoft Excel Com...	1KB
students2	2020-04-05 오후 5:52	텍스트 문서	1KB
students3	2020-04-05 오후 5:53	텍스트 문서	1KB

Console Output:

```
..$ sp.mobile
..$ sp.device
..$ sp.online.co
..$ sp.offline.co
..$ year
$ : 'data.frame':
..$ pid
..$ age
..$ gender
..$ Mobile.lv
..$ Telecom
..$ Smartphone.br
..$ sp.mobile
..$ sp.device
..$ sp.online.co
..$ sp.offline.co
..$ year
$ : 'data.frame':
..$ pid
..$ Smartphone.time: num [1:10] 0 60 345 90 40 65 170 95 85 75
..$ SNS.time : num [1:10] 0 0 10 0 0 0 0 0 0
..$ year : num [1:10] 2019 2019 2019 2019 2019 ...
> save(List.KMP, file="List_KMP.Rdata")
>
```

Environment Panel:

Size	Modified
339.5 KB	Apr 9, 2020, 12:40 AM
23 KB	Apr 15, 2020, 6:07 AM
218 B	Apr 15, 2020, 9:11 PM
37.7 KB	Mar 25, 2020, 12:14 PM
18.7 KB	Apr 9, 2020, 9:28 PM
398.4 KB	Apr 7, 2020, 9:03 PM
37.2 KB	Apr 15, 2020, 9:11 PM
37.7 KB	Mar 24, 2020, 1:50 AM
572 B	Apr 8, 2020, 8:32 PM
465 B	Apr 8, 2020, 8:32 PM
114 B	Apr 5, 2020, 5:52 PM
112 B	Apr 5, 2020, 5:51 PM
116 B	Apr 5, 2020, 5:52 PM
114 B	Apr 5, 2020, 5:52 PM
130 B	Apr 5, 2020, 5:53 PM

Text Overlay:

Copy List_KMP.Rdata and Paste the file into this folder

Load List.KMP

The screenshot displays the RStudio interface with the following components:

- Source Editor:** Contains R code for data filtering and summarization. The code filters data for the years 2018 and 2019, summarizes variables (mean, na.rm=T), and saves the resulting list as 'List.KMP.RData'.
- Environment:** Shows the 'Global Environment' with a list of objects including 'List.KMP' (List of 4), 'matched_ctry_names' (184 obs. of 4 variables), and several 'p' objects (e.g., 'p17', 'p17.df', 'p17.s', 'p1789', 'p18', 'p18.s', 'p19') representing different data frames.
- Files:** A file explorer view showing a directory of files, including 'List.KMP.RData' which is circled in orange. Other files include various '.dta' and '.csv' files.
- Console:** Shows the output of the R code, including summary statistics for 'sp.mobile', 'sp.device', 'sp.online.content', 'sp.offline.contet', and 'year' for both 2018 and 2019 data frames.

Simply click the file

OK, let's see the structure of List.KMP

Str(List.KMP)

```
List of 4
$ : 'data.frame':      10 obs. of  11 variables:
..$ pid                : num [1:10] 9920004 12500003 17350004 17670001 23860001 ...
..$ age                : num [1:10] 9 26 12 55 70 58 50 68 39 37
..$ gender              : Factor w/ 2 levels "Male","Female": 1 1 2 1 1 1 1 1 1 2
..$ Mobile.lv          : Factor w/ 5 levels "3G","LTE","LTE-A",...: 5 2 2 2 5 2 2 2 2 3
..$ Telecom            : Factor w/ 5 levels "SKT","KT","LG U+",...: 5 2 3 1 5 1 1 1 2 2
..$ Smartphone.brand   : Factor w/ 8 levels "Samsung","Apple",...: 8 1 1 1 8 1 1 1 1 3
..$ sp.mobile          : num [1:10] 0 42 19 38 18 65 32 58 63 54
..$ sp.device          : num [1:10] 0 0 0 0 0 0 0 10 0 10
..$ sp.online.content: num [1:10] 0 53 19 38 18 65 32 68 63 114
..$ sp.offline.contet: num [1:10] 10 22 0 0 0 50 36 0 25 40
..$ year               : num [1:10] 2017 2017 2017 2017 2017 ...
$ : 'data.frame':      10 obs. of  11 variables:
..$ pid                : num [1:10] 9920004 12500003 17350004 17670001 23860001 ...
..$ age                : num [1:10] 10 27 13 56 71 59 51 69 40 38
..$ gender              : Factor w/ 2 levels "Male","Female": 1 1 2 1 1 1 1 1 1 2
..$ Mobile.lv          : Factor w/ 5 levels "3G","LTE","LTE-A",...: 5 3 3 2 5 2 2 2 3 3
..$ Telecom            : Factor w/ 5 levels "SKT","KT","LG U+",...: 5 3 3 2 5 3 1 2 1 2
..$ Smartphone.brand   : Factor w/ 8 levels "Samsung","Apple",...: 8 1 3 1 8 3 1 3 1 3
..$ sp.mobile          : num [1:10] 0 90 20 39 30 80 33 36 40 59
..$ sp.device          : num [1:10] 0 60 35 0 0 15 0 10 10 12
..$ sp.online.content: num [1:10] 0 359 55 39 30 95 33 46 50 71
..$ sp.offline.contet: num [1:10] 8 120 0 0 0 12 20 0 0 100
..$ year               : num [1:10] 2018 2018 2018 2018 2018 ...
$ : 'data.frame':      10 obs. of  11 variables:
..$ pid                : num [1:10] 9920004 12500003 17350004 17670001 23860001 ...
..$ age                : num [1:10] 11 28 14 57 72 60 52 70 41 39
..$ gender              : Factor w/ 2 levels "Male","Female": 1 1 2 1 1 1 1 1 1 2
..$ Mobile.lv          : Factor w/ 5 levels "3G","LTE","LTE-A",...: 5 2 3 2 2 2 2 2 3 3
..$ Telecom            : Factor w/ 5 levels "SKT","KT","LG U+",...: 5 3 3 1 2 1 1 2 3 3
..$ Smartphone.brand   : Factor w/ 8 levels "Samsung","Apple",...: 8 2 1 1 1 1 1 3 3 3
..$ sp.mobile          : num [1:10] 0 60 21 30 32 78 35 54 73 45
..$ sp.device          : num [1:10] 0 30 0 0 0 0 0 20 15
..$ sp.online.content: num [1:10] 0 90 21 30 32 78 35 54 93 60
..$ sp.offline.contet: num [1:10] 21 0 0 20 0 24 90 0 20 60
..$ year               : num [1:10] 2019 2019 2019 2019 2019 ...
$ : 'data.frame':      10 obs. of  4 variables:
..$ pid                : num [1:10] 9920004 12500003 17350004 17670001 23860001 ...
..$ Smarphone.time: num [1:10] 0 60 345 90 40 65 170 95 85 75
..$ SNS.time          : num [1:10] 0 0 10 0 0 0 0 0 0 0
..$ year               : num [1:10] 2019 2019 2019 2019 2019 ...
```

Let's give names for the list elements

```
names(List.KMP)
```

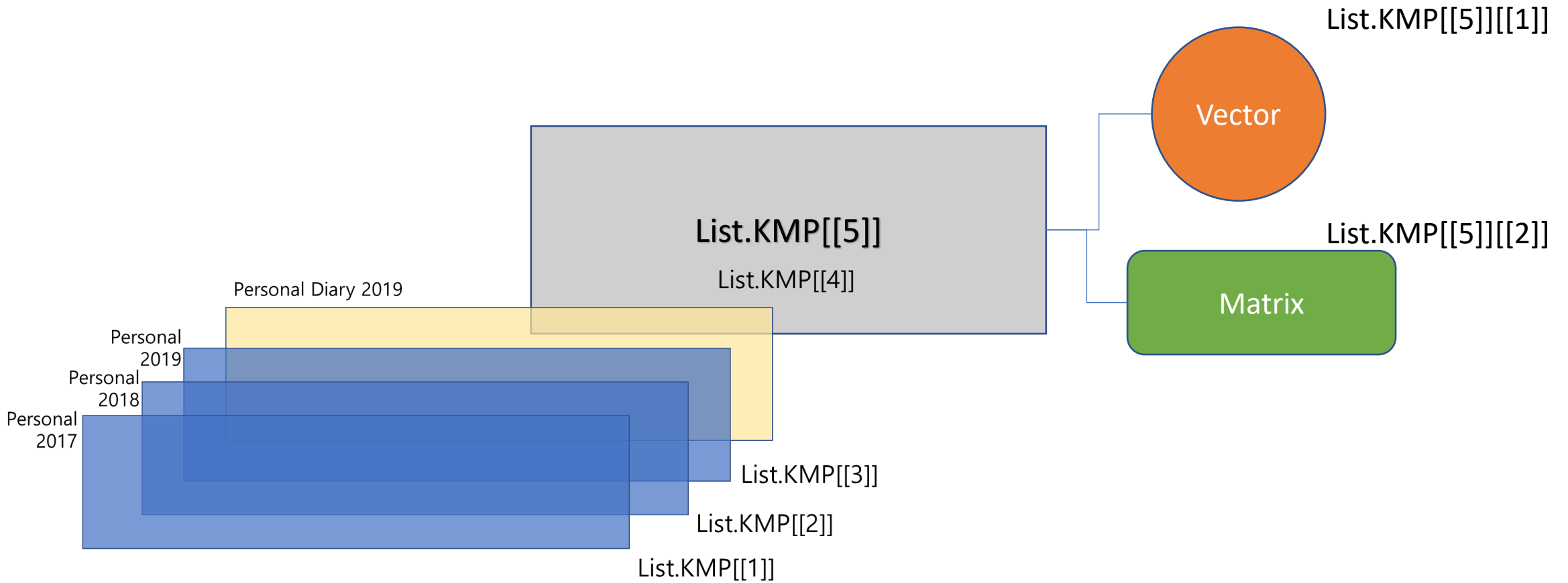
```
names(List.KMP) <- c("p17", "p18", "p19", "d19")
```

How can we extract the first element?

```
List.KMP[[1]]
```

```
List.KMP[["p17"]]
```

List of List??

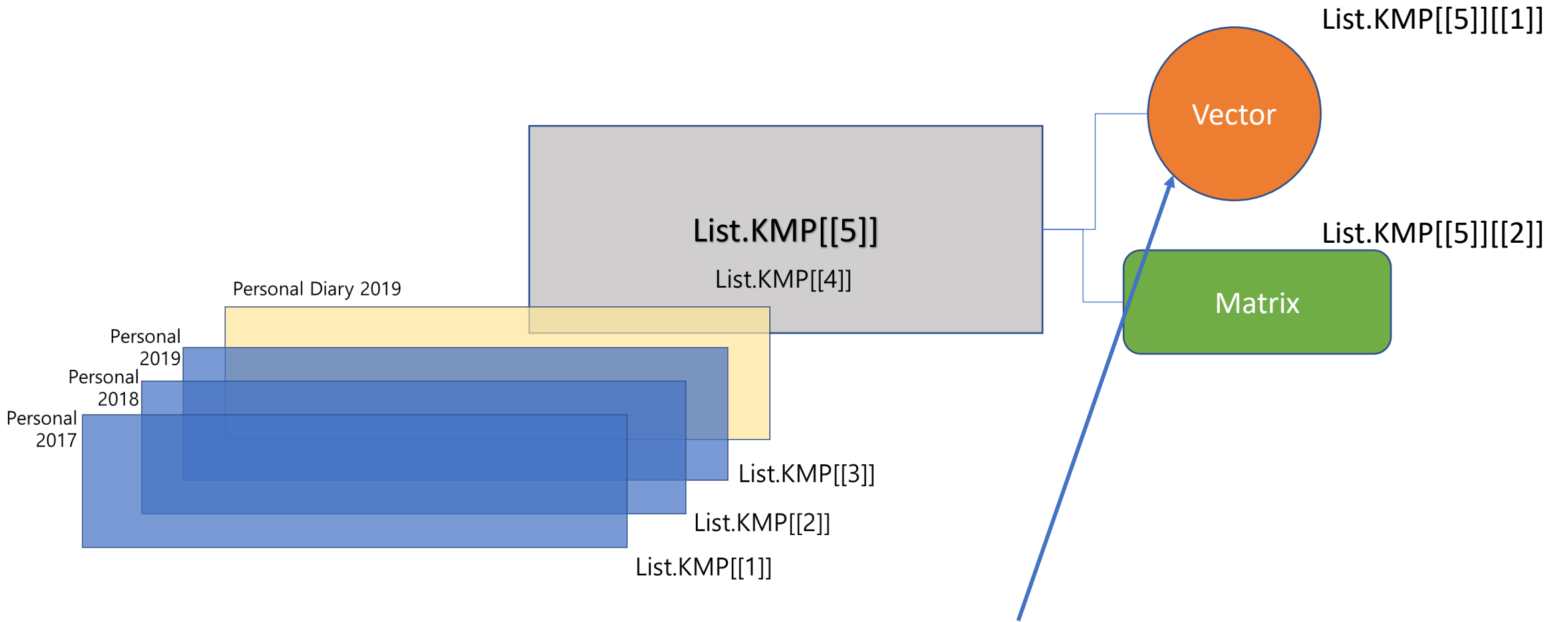


```
List.KMP[[5]]<-list(0)
```

```
List.KMP[[5]][[1]]<-c(1:10)
```

```
List.KMP[[5]][[2]]<-matrix(c(1:12), nrow=4)
```

List of List??



How to extract this element from List.KMP?

Let's draw only p17

```
p17.df<-List.KMP[["p17"]]
```

Summary statistics

```
summary(p17.df)
```

Among 10 people,
How many people use Samsung phone?

How much do people spend for the mobile communication **on average?**

Do the same thing to 2019 dataset and answer the questions below.

1. In 2019, how many people use Samsung phone?
2. Draw boxplot of people's spending on the mobile communication
*Hint: use `boxplot()`

Do the same thing to 2019 dataset and answer the questions below.

1. In 2019, how many people use Samsung phone?

```
> summary(p19.df)
  pid          age      gender      Mobile.Tv      Telecom      Smartphone.brand
Min.   : 9920004  Min.   :11.00  Male   :8    3G           :0    SKT           :3    Samsung       :5
1st Qu.:17430003 1st Qu.:30.75  Female:2    LTE          :6    KT            :2    LG             :3
Median :24155001 Median :46.50          LTE-A        :3    LG U+         :4    Apple          :1
Mean    :31235002 Mean    :44.40          5G           :0    MVNO          :0    No Smartphone:1
3rd Qu.:47107501 3rd Qu.:59.25          No Smartphone:1 No Smartphone:1 Pantech        :0
Max.    :65840001 Max.    :72.00          (Other)      :0

  sp.mobile      sp.device      sp.online.content      sp.offline.contet      year
Min.   : 0.0      Min.   : 0.00      Min.   : 0.0      Min.   : 0.00      Min.   :2019
1st Qu.:30.5      1st Qu.: 0.00      1st Qu.:30.5      1st Qu.: 0.00      1st Qu.:2019
Median :40.0      Median : 0.00      Median :44.5      Median :20.00      Median :2019
Mean    :42.8      Mean    : 6.50      Mean    :49.3      Mean    :23.50      Mean    :2019
3rd Qu.:58.5      3rd Qu.:11.25      3rd Qu.:73.5      3rd Qu.:23.25      3rd Qu.:2019
Max.    :78.0      Max.    :30.00      Max.    :93.0      Max.    :90.00      Max.    :2019
```

In 2019 dataset,

1. Draw histogram of people's spending on the mobile communication *Hint: use hist()

2. Try this code and explain the role of the option

```
hist(p19.df$sp.mobile, breaks=10)
```

3. Try this code and explain the role of the option

```
hist(p19.df$sp.mobile, breaks=10, freq=T)
```

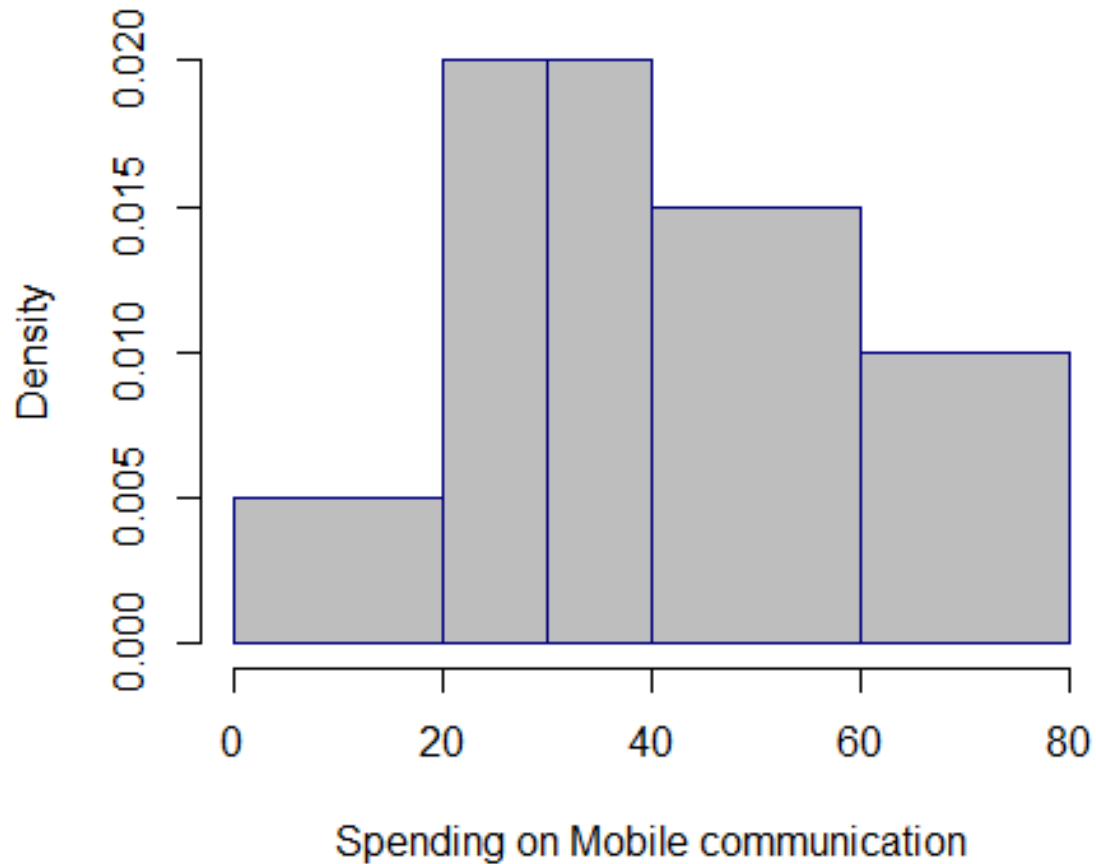
4. Try this code and explain the role of the option

```
hist(p19.df$sp.mobile, breaks=c(0,20,40,100))
```

5. Try this code and explain the role of the option

```
hist(p19.df$sp.mobile, las=1)
```

Histogram of spending on mobile com



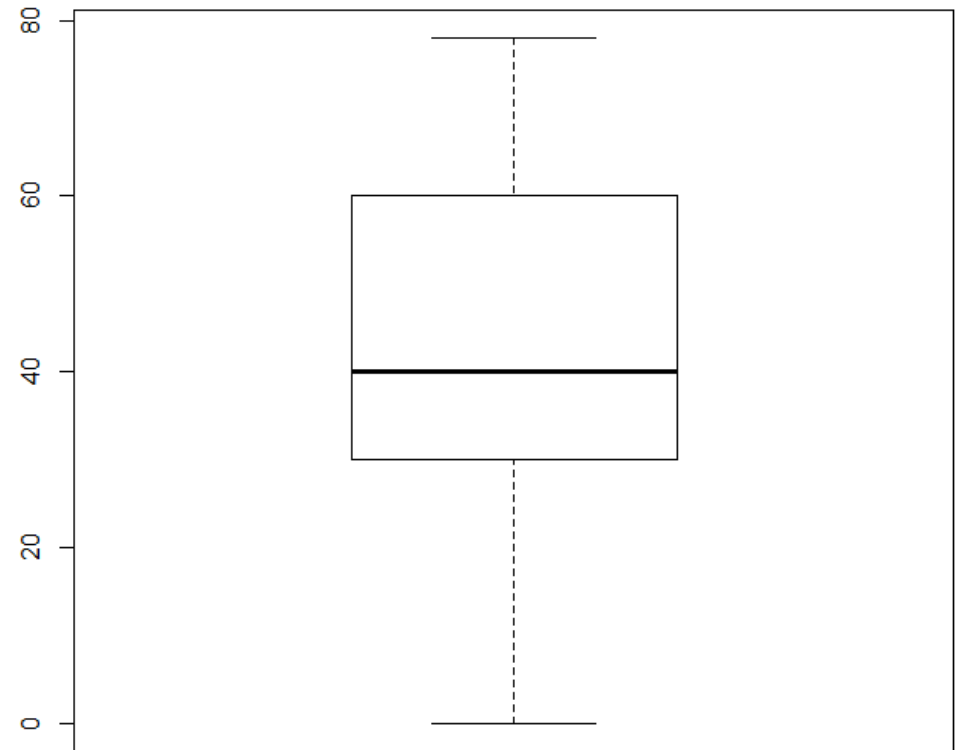
Make the graph left

- col="filled color"
- border="border color"
- main="Main title"
- xlab="x-axis label"
- breaks=c()

In 2019 dataset,

1. Draw boxplot of people's spending on the mobile communication

*Hint: use `boxplot()`



Did 'KT' users in 2017 change their telecom companies in 2018 and 2019?

1. Let's extract data.frames from the List

```
p17.df<-List.KMP[["p17"]]  
p18.df<-List.KMP[["p18"]]
```

2. Find the pid(personal ID) whose telecom company was "KT"

Hint: `Vector.name <- p17.df[== "KT", " "]`



3. Try the code below and explain how it works

```
kt.user.18<-p18.df[p18.df$pid %in% kt.user.17.pid, c("pid", "Telecom")]  
kt.user.19<-p19.df[p19.df$pid %in% kt.user.17.pid, c("pid", "Telecom")]
```

4. Merge the two data.frames above by the key variable "pid", and Change the second and the third column names to "y2018" and "y2019"

Hint 1: `DF3<-merge(DF1, DF2, by="pid")`

Hint 2: `name(DF3)[c(,)]<-c(" ", " ")`

Did 'Samsung' users in 2017 change their phone brand in 2018 and 2019?