

Show
me the
data!

Week08: Time Series Analysis

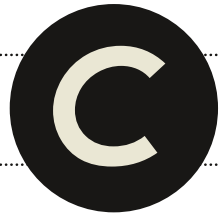
Big Data & Social Analysis R

Instructors: Chung-pei Pien

ZU1942001/266868001/Z23937001/ZM1941001



International College of
INNOVATION
National Chengchi University
國立政治大學創新國際學院



CONTENTS



1 Time Series Analysis

Time Series Analysis

Time Series data is data that is observed at a fixed interval time and it could be measured daily, monthly, annually, etc (Khalid 2020).

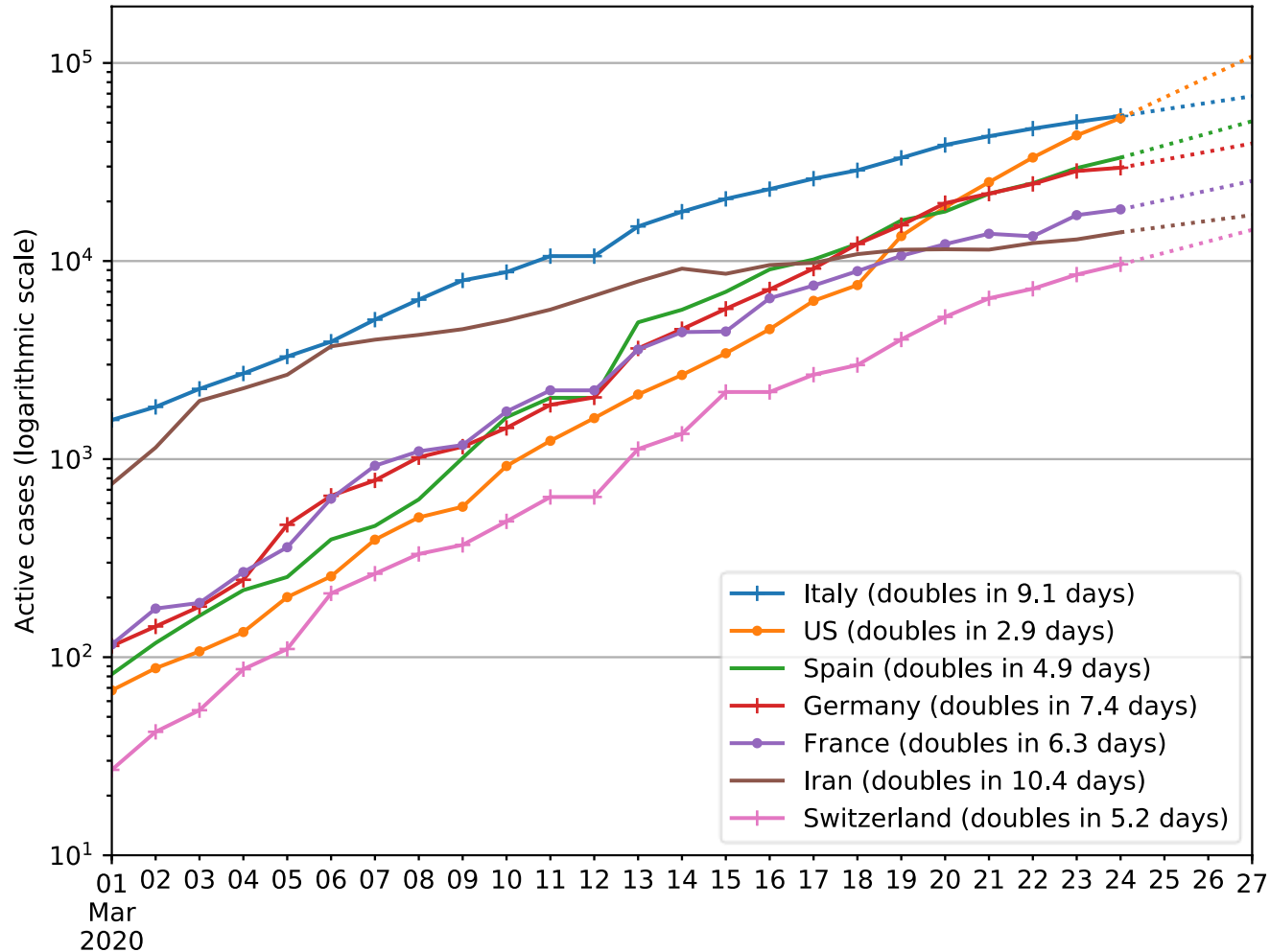


01

Time Series Analysis

Introduction

Active COVID-19 cases in the most affected countries
as of 2020-03-24, including presumptive positive cases



We focused on two kinds of an object: numerics or characters.

If you attempt to employ time series analysis, you have to transform an object into time format.

as.Date

as.POSIXct

	Year	Month	Day	Hour	Min.	Sec.	Time Zone
as.Date	●	●	●				●
as.POSIXct	●	●	●	●	●	●	●

```
date_1 <- as.Date("2015-4-7")
```

```
date_2 <- as.Date("2015-4-7 10:15")
```

```
date_3 <- as.Date("2015/4/07 10:15")
```


Read `trumptweets.xlsx`.

Create a plot to show the daily number of Trump tweets from June 1, 2017 to June 30, 2017.

01

Time Series Analysis

Trump Tweets

```
trumptweets <- read_xlsx("trumptweets.xlsx")
```

```
trumptweets$date <-  
  as.Date(trumptweets$created_at)
```

created_at
2018-05-18 20:41:21
2018-05-18 20:22:25
2018-05-18 20:02:49
2018-05-18 16:34:24
2018-05-18 15:05:35
2018-05-18 13:50:11
2018-05-18 13:38:11
2018-05-18 10:51:43
2018-05-18 10:24:53
2018-05-17 23:44:23

date
2018-05-18
2018-05-18
2018-05-18
2018-05-18
2018-05-18
2018-05-18
2018-05-18
2018-05-18
2018-05-18
2018-05-17

```
trump_june17 <- trumptweets %>%
  filter(date >= "2017-06-01" &
         date <= "2017-06-30")
```

	status_id	created_at	user_id	screen_name	text	source
1	880919871353217024	2017-06-30 22:43:58	25073877	realDonaldTrump	America will THINK BIG once again. We will inspire millions ...	Twitter Ads
2	880892632142143490	2017-06-30 20:55:44	25073877	realDonaldTrump	The era of strategic patience with the North Korea regime h...	Twitter for iPhone
3	880853246126022656	2017-06-30 18:19:13	25073877	realDonaldTrump	I am encouraged by President Moon's assurances that he wi...	Twitter Web Client
4	880771685460344832	2017-06-30 12:55:08	25073877	realDonaldTrump	Watched low rated @Morning_Joe for first time in long time...	Twitter for iPhone
5	880739852286918656	2017-06-30 10:48:38	25073877	realDonaldTrump	Crime and killings in Chicago have reached such epidemic p...	Twitter for iPhone
6	880737163247267840	2017-06-30 10:37:57	25073877	realDonaldTrump	If Republican Senators are unable to pass what they are wor...	Twitter for iPhone
7	880602793341591552	2017-06-30 01:44:01	25073877	realDonaldTrump	Just finished a very good meeting with the President of Sou...	Twitter for iPhone
8	880564232588107779	2017-06-29 23:10:47	25073877	realDonaldTrump	Our new American Energy Policy will unlock MILLIONS of jo...	Twitter Web Client
9	880563788705148928	2017-06-29 23:09:01	25073877	realDonaldTrump	When it comes to the future of America's energy needs, we ...	Twitter for iPhone
10	880559961855729664	2017-06-29 22:53:49	25073877	realDonaldTrump	Statement on House Passage of Kate's Law and No Sanctuar...	Twitter for iPhone
11	880553561008078848	2017-06-29 22:28:23	25073877	realDonaldTrump	Good news out of the House with the passing of 'No Sanctu...	Twitter for iPhone
12	880540637358981120	2017-06-29 21:37:02	25073877	realDonaldTrump	Good news, House just passed #KatesLaw. Hopefully Senate...	Twitter for iPhone
13	880534684479897600	2017-06-29 21:13:22	25073877	realDonaldTrump	"Mattis Says Trump's Warning Stopped Chemical Weapons ...	Twitter for iPhone
14	880508977590697984	2017-06-29 19:31:13	25073877	realDonaldTrump	Join me LIVE with @VP, @SecretaryPerry, @SecretaryZinke a...	Twitter for iPhone

01

Time Series Analysis

Trump Tweets

```
trump_date <- trump_june17 %>%  
  group_by(date) %>%  
  summarise(post = n())
```

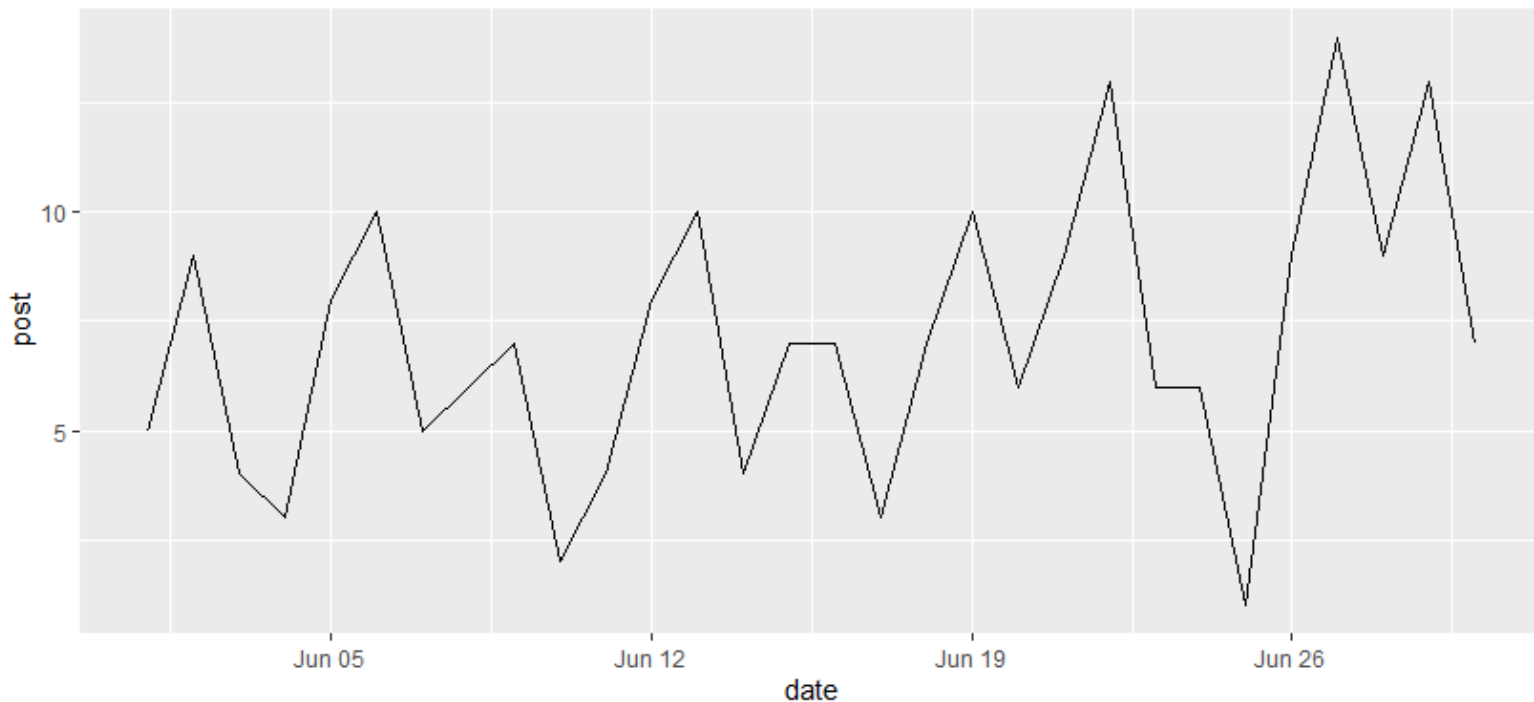
	date	post
1	2017-06-01	5
2	2017-06-02	9
3	2017-06-03	4
4	2017-06-04	3
5	2017-06-05	8
6	2017-06-06	10
7	2017-06-07	5
8	2017-06-09	7
9	2017-06-10	2
10	2017-06-11	4
11	2017-06-12	8
12	2017-06-13	10
13	2017-06-14	4
14	2017-06-15	7
15	2017-06-16	7
16	2017-06-17	3
17	2017-06-18	7
18	2017-06-19	10

01

Time Series Analysis

Trump Tweets

```
ggplot(trump_date, aes(x = date, y = post)) +  
  geom_line()
```



Read `google_mobility_data.csv`

On May 15th, 2021, Taiwan CDC has raised epidemic warning to Level 3 in Taipei and New Taipei

**Taipei, New Taipei
Level 3
COVID-19 Alert**

**May
15-28**

- Work and school continue under CECC guidelines
- Wear a mask at all times unless at home
- Closure of movie theaters, sports centers, libraries, exhibitions, bars, and adult entertainment venues
- Gathering of 5+ persons indoors or 10+ outdoors prohibited
- Restaurants use ID-based registration system, partitions for dine-in (or) carry-out/delivery only
- Avoid unnecessary travel between Level 2 and Level 3 alert zones

Source: CECC / CNA graphic

Level 3 in Taiwan can be treated as soft lockdown.

Does the soft-lockdown order have influences on Taiwanese mobility?

Create a plot to show Taiwan's transit.stations' changes from Jan. 1, 2021 to May 18, 2021

Practice 1

However, we usually get diverse time data formats.

created_at
2018-05-18 20:41:21
2018-05-18 20:22:25
2018-05-18 20:02:49
2018-05-18 16:34:24
2018-05-18 15:05:35
2018-05-18 13:50:11
2018-05-18 13:38:11
2018-05-18 10:51:43
2018-05-18 10:24:53
2018-05-17 23:44:23

We are very lucky that Google Mobility and Trump tweets datasets provide default R time format.


```
date_4 <- as.Date("20150407")
```

In this case, we need to tell R that the structure of the time data.

```
date_4 <- as.Date("20150407", "%Y%m%d")
```

```
date_5 <- as.Date("150407", "%y%m%d")
```

```
date_6 <- as.Date("11022012")
```

```
date_6 <- as.Date("11022012", "%m%d%Y")
```

```
date_7 <- as.Date("11-02-2012")
```

```
date_7 <- as.Date("11-02-2012", "%m%d%Y")
```

```
date_7 <- as.Date("11-02-2012", "%m-%d-%Y")
```

01

Time Series Analysis

Date and Time Format

```
date_8 <- as.Date("Dec022012")
```

```
date_8 <- as.Date("Dec022012", "%m%d%Y")
```

```
date_8 <- as.Date("Dec022012", "%b%d%Y")
```

date_6	2012-11-02 UTC
date_7	2012-11-02 UTC
date_8	

Why %b does not work?

Because R only recognizes abbreviated month name in the language your OS uses.

```
format(Sys.Date(), "%b")
```

```
> format(Sys.Date(), "%b")  
[1] "四月"
```

01

Time Series Analysis

Date and Time Format

```
Sys.setlocale(locale = "English")
```

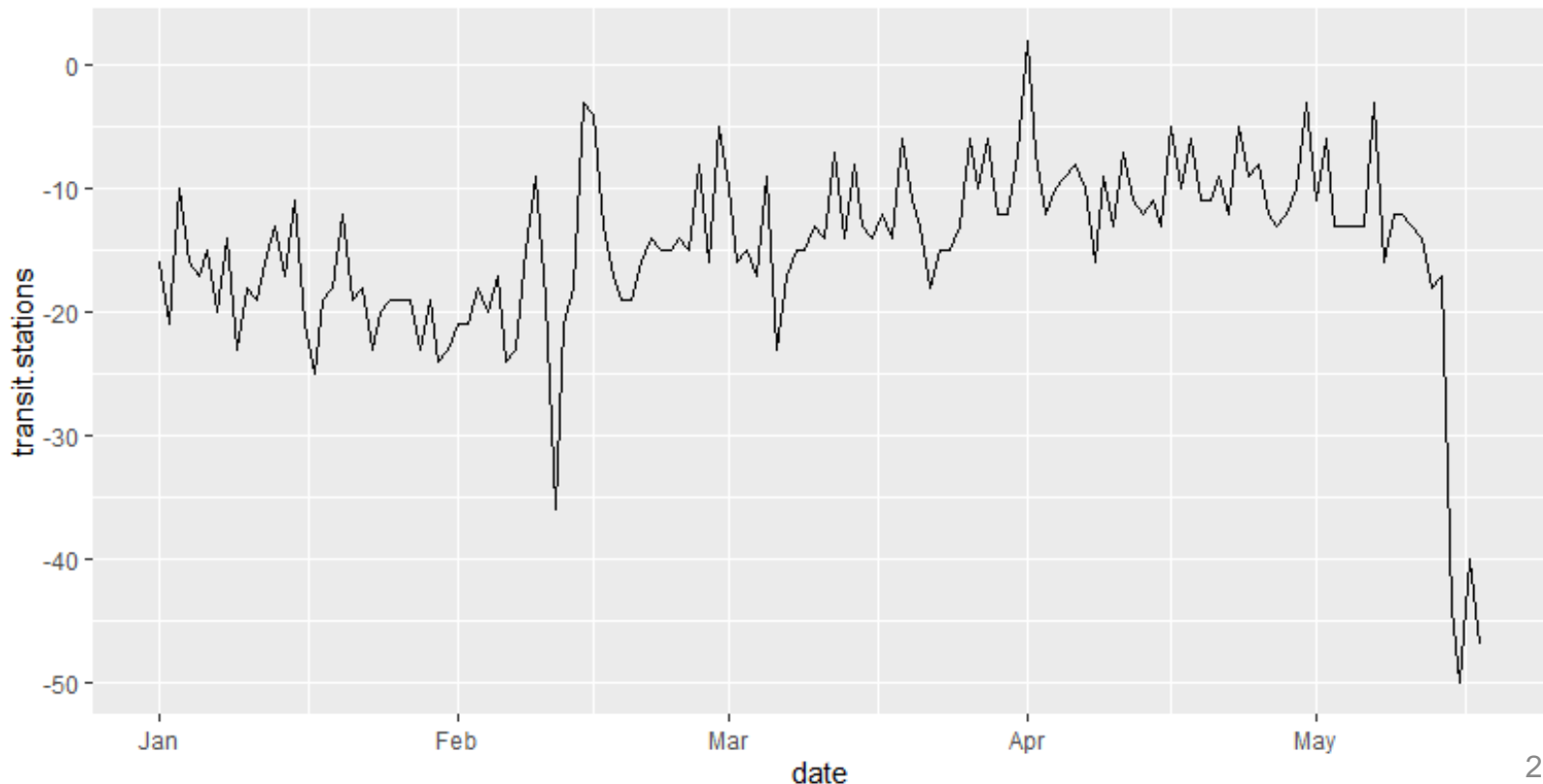
```
date_8 <- as.Date("Dec022012", "%b%d%Y")
```

01

Time Series Analysis

Advanced Dataframe Reshaping

The plot we created earlier just show Taiwan's transit.stations' change from Jan. 1, 2021 to May 18, 2021



01

Time Series Analysis

Advanced Dataframe Reshaping

We want to compare all mobility ways' changes



The dataframe that ggplot2 can draw multilines in a plot.

date	move	value
2021-01-01	retail.and.recreation	2
2021-01-01	grocery.and.pharmacy	7
2021-01-01	parks	34
2021-01-01	transit.stations	-16
2021-01-01	workplaces	-61
2021-01-01	residential	14
2021-01-02	retail.and.recreation	-11
2021-01-02	grocery.and.pharmacy	-2
2021-01-02	parks	4
2021-01-02	transit.stations	-21
2021-01-02	workplaces	-14
2021-01-02	residential	4

01

Time Series Analysis

Advanced Dataframe Reshaping

However, our original dataset is like this:

date	retail.and.recreation	grocery.and.pharmacy	parks	transit.stations	workplaces	residential
2021-01-01	2	7	34	-16	-61	14
2021-01-02	-11	-2	4	-21	-14	4
2021-01-03	-7	-1	17	-10	-8	1
2021-01-04	-17	-4	-26	-16	2	3
2021-01-05	-17	0	-26	-17	6	2
2021-01-06	-18	-3	-27	-15	6	2
2021-01-07	-20	-9	-43	-20	3	0
2021-01-08	-17	-3	-39	-14	0	4
2021-01-09	-12	3	-31	-23	2	4
2021-01-10	-5	5	-8	-18	-8	3
2021-01-11	-20	-4	-42	-19	-1	4
2021-01-12	-13	6	-29	-16	3	3
2021-01-13	-13	2	-17	-13	4	1
2021-01-14	-13	-5	-23	-17	0	-3
2021-01-15	-11	-4	-18	-11	-4	-1
2021-01-16	-14	-5	-18	-21	-4	0
2021-01-17	-14	-5	-18	-25	-15	2



date	move	value
2021-01-01	retail.and.recreation	2
2021-01-01	grocery.and.pharmacy	7
2021-01-01	parks	34
2021-01-01	transit.stations	-16
2021-01-01	workplaces	-61
2021-01-01	residential	14
2021-01-02	retail.and.recreation	-11
2021-01-02	grocery.and.pharmacy	-2
2021-01-02	parks	4
2021-01-02	transit.stations	-21
2021-01-02	workplaces	-14
2021-01-02	residential	4

```
install.packages("tidyr")  
library(tidyr)
```

```
gather() and spread()
```

01

Time Series Analysis

Advanced Dataframe Reshaping

country	year	cases
Afghanistan	1999	745
Afghanistan	2000	2666
Brazil	1999	37737
Brazil	2000	80488
China	1999	212258
China	2000	213766

country	1999	2000
Afghanistan	745	2666
Brazil	37737	80488
China	212258	213766

`g_w`

```
g_l <- g_w %>%
  gather("year", "cases", 2:3)
```

Which columns you
want to gather

A new column name to
store `g_w`'s column names

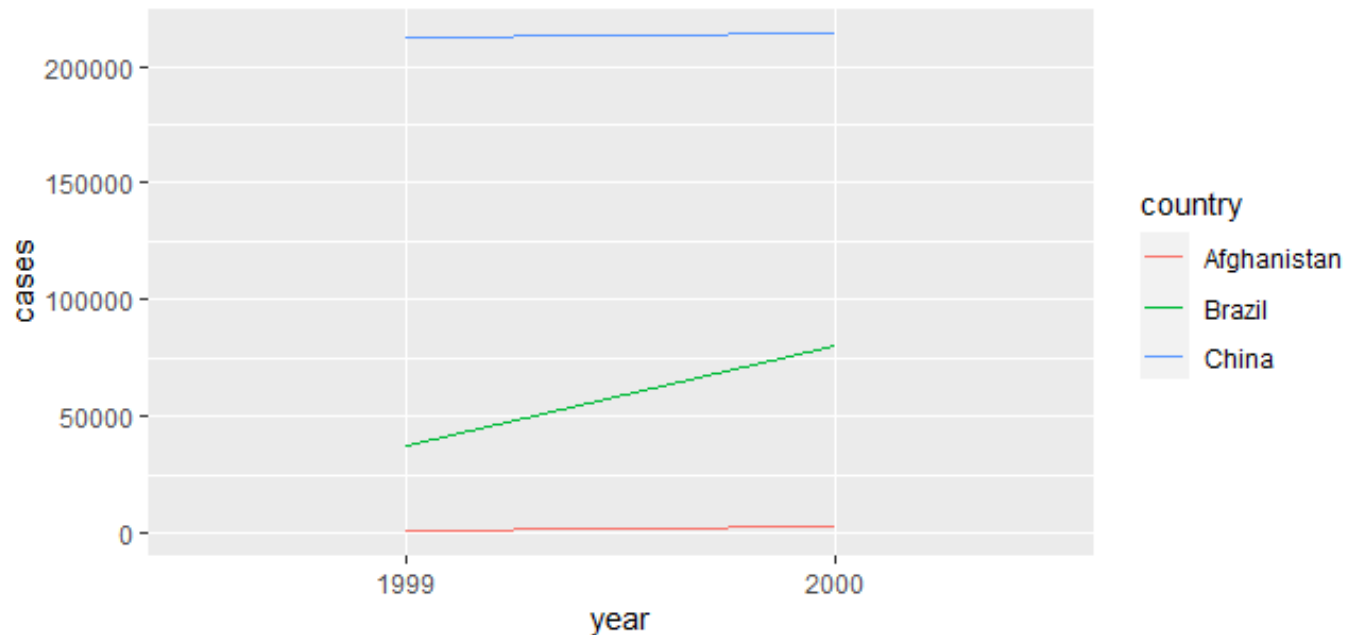
A new column name to store
the values of `g_w`'s columns

01

Time Series Analysis

Advanced Dataframe Reshaping

```
ggplot(g_l, aes(x = year, y = cases, color = country,  
group = country)) +  
  geom_line()
```



01

Time Series Analysis

Advanced Dataframe Reshaping

country	year	key	value
Afghanistan	1999	cases	745
Afghanistan	1999	population	19987071
Afghanistan	2000	cases	2666
Afghanistan	2000	population	20595360
Brazil	1999	cases	37737
Brazil	1999	population	172006362
Brazil	2000	cases	80488
Brazil	2000	population	174504898
China	1999	cases	212258
China	1999	population	1272915272
China	2000	cases	213766
China	2000	population	1280428583

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

```
s_w <- s_l %>%
  spread(key, value)
```

s_l

s_l's column name' contents
become new column names.

s_l's column name that is value.

01

Time Series Analysis

Advanced Dataframe Reshaping

date	retail.and.recreation	grocery.and.pharmacy	parks	transit.stations	workplaces	residential
2021-01-01	2	7	34	-16	-61	14
2021-01-02	-11	-2	4	-21	-14	4
2021-01-03	-7	-1	17	-10	-8	1
2021-01-04	-17	-4	-26	-16	2	3
2021-01-05	-17	0	-26	-17	6	2
2021-01-06	-18	-3	-27	-15	6	2
2021-01-07	-20	-9	-43	-20	3	0
2021-01-08	-17	-3	-39	-14	0	4
2021-01-09	-12	3	-31	-23	2	4
2021-01-10	-5	5	-8	-18	-8	3
2021-01-11	-20	-4	-42	-19	-1	4
2021-01-12	-13	6	-29	-16	3	3
2021-01-13	-13	2	-17	-13	4	1
2021-01-14	-13	-5	-23	-17	0	-3
2021-01-15	-11	-4	-18	-11	-4	-1
2021-01-16	-14	-5	-18	-21	-4	0
2021-01-17	-14	-5	-18	-25	-15	2



date	move	value
2021-01-01	retail.and.recreation	2
2021-01-01	grocery.and.pharmacy	7
2021-01-01	parks	34
2021-01-01	transit.stations	-16
2021-01-01	workplaces	-61
2021-01-01	residential	14
2021-01-02	retail.and.recreation	-11
2021-01-02	grocery.and.pharmacy	-2
2021-01-02	parks	4
2021-01-02	transit.stations	-21
2021-01-02	workplaces	-14
2021-01-02	residential	4

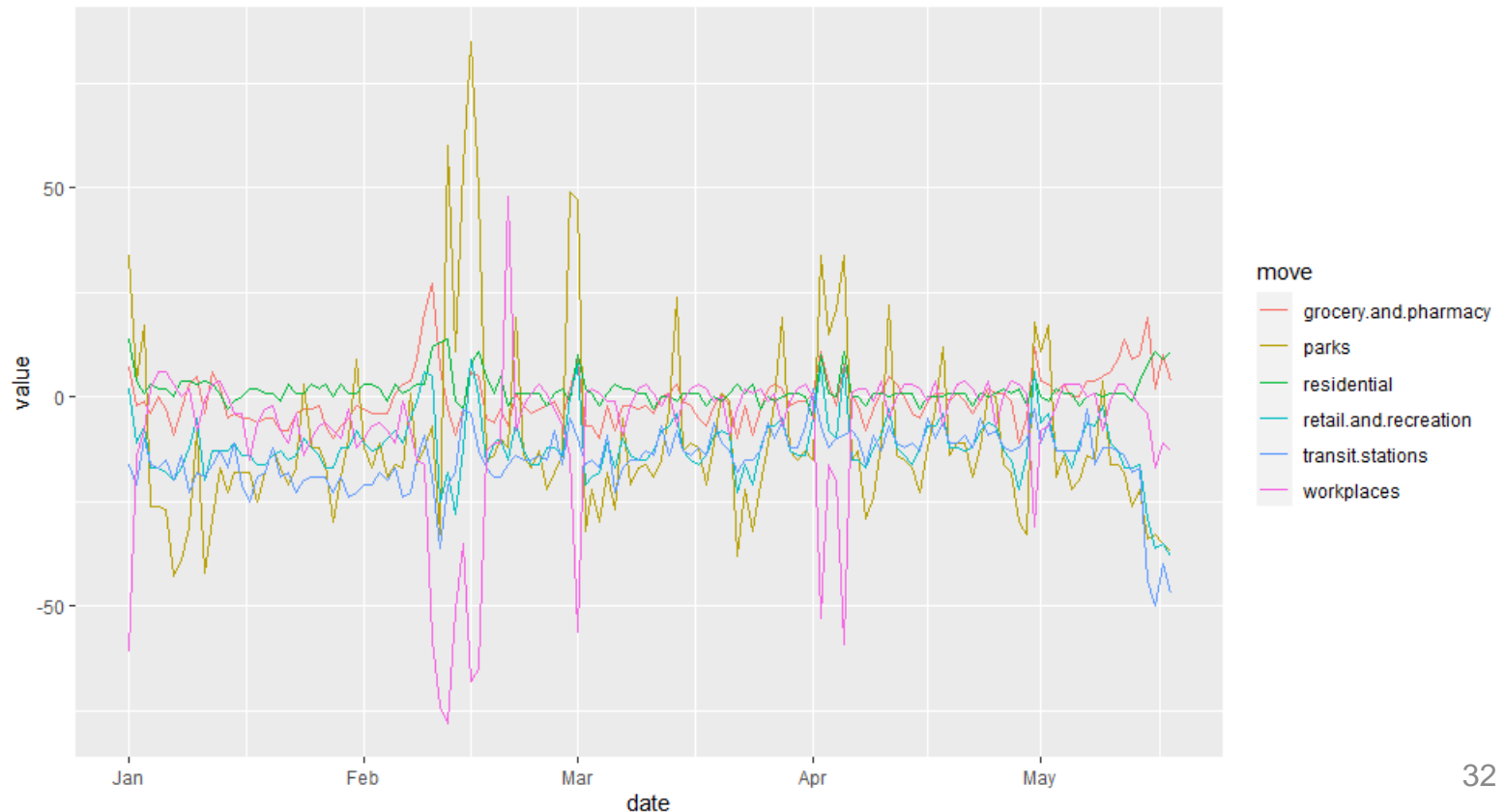
```
tw_google_l <- tw_google[, -c(1:4)] %>%
  gather("move", "value", 2:7)
```

01

Time Series Analysis

Advanced Dataframe Reshaping

```
ggplot(tw_google_l, aes(x = date, y = value,  
  group = move, color = move)) +  
  geom_line()
```




```
time_date_1 <- as.POSIXct("2015-04-07 10:15:30")
```

```
time_date_2 <- as.POSIXct("2015-04-07 101530")
```

```
time_date_3 <- as.POSIXct("20150407 101530")
```

```
time_date_3 <- as.POSIXct("20150407 101530", "%Y%m%d %H%M%S")
```

```
Sys.timezone()
```

```
time_date_4 <- as.POSIXct("2015-04-07 10:15:30", tz = "UTC")
```

```
time_date_1 - time_date_4
```

Create a plot to show the hourly number of Trump tweets.

	status_id	created_at	user_id	screen_name	text	source
1	880919871353217024	2017-06-30 22:43:58	25073877	realDonaldTrump	America will THINK BIG once again. We will inspire millions ...	Twitter Ads
2	880892632142143490	2017-06-30 20:55:44	25073877	realDonaldTrump	The era of strategic patience with the North Korea regime h...	Twitter for iPhone
3	880853246126022656	2017-06-30 18:19:13	25073877	realDonaldTrump	I am encouraged by President Moon's assurances that he wi...	Twitter Web Client
4	880771685460344832	2017-06-30 12:55:08	25073877	realDonaldTrump	Watched low rated @Morning_Joe for first time in long time...	Twitter for iPhone
5	880739852286918656	2017-06-30 10:48:38	25073877	realDonaldTrump	Crime and killings in Chicago have reached such epidemic p...	Twitter for iPhone
6	880737163247267840	2017-06-30 10:37:57	25073877	realDonaldTrump	If Republican Senators are unable to pass what they are wor...	Twitter for iPhone
7	880602793341591552	2017-06-30 01:44:01	25073877	realDonaldTrump	Just finished a very good meeting with the President of Sou...	Twitter for iPhone
8	880564232588107779	2017-06-29 23:10:47	25073877	realDonaldTrump	Our new American Energy Policy will unlock MILLIONS of jo...	Twitter Web Client
9	880563788705148928	2017-06-29 23:09:01	25073877	realDonaldTrump	When it comes to the future of America's energy needs, we ...	Twitter for iPhone
10	880559961855729664	2017-06-29 22:53:49	25073877	realDonaldTrump	Statement on House Passage of Kate's Law and No Sanctuar...	Twitter for iPhone
11	880553561008078848	2017-06-29 22:28:23	25073877	realDonaldTrump	Good news out of the House with the passing of 'No Sanctu...	Twitter for iPhone
12	880540637358981120	2017-06-29 21:37:02	25073877	realDonaldTrump	Good news, House just passed #KatesLaw. Hopefully Senate...	Twitter for iPhone
13	880534684479897600	2017-06-29 21:13:22	25073877	realDonaldTrump	"Mattis Says Trump's Warning Stopped Chemical Weapons ...	Twitter for iPhone
14	880508977590697984	2017-06-29 19:31:13	25073877	realDonaldTrump	Join me LIVE with @VP, @SecretaryPerry, @SecretaryZinke a...	Twitter for iPhone

01

Time Series Analysis

Trump Tweets

```
trumptweets$time <-  
  as.POSIXct(trumptweets$created_at)
```

created_at
2018-05-18 20:41:21
2018-05-18 20:22:25
2018-05-18 20:02:49
2018-05-18 16:34:24
2018-05-18 15:05:35
2018-05-18 13:50:11
2018-05-18 13:38:11
2018-05-18 10:51:43
2018-05-18 10:24:53
2018-05-17 23:44:23

time
2018-05-18 20:41:21
2018-05-18 20:22:25
2018-05-18 20:02:49
2018-05-18 16:34:24
2018-05-18 15:05:35
2018-05-18 13:50:11
2018-05-18 13:38:11
2018-05-18 10:51:43
2018-05-18 10:24:53
2018-05-17 23:44:23

```
trumptweets$hour <-  
  format(trumptweets$time, "%H")
```

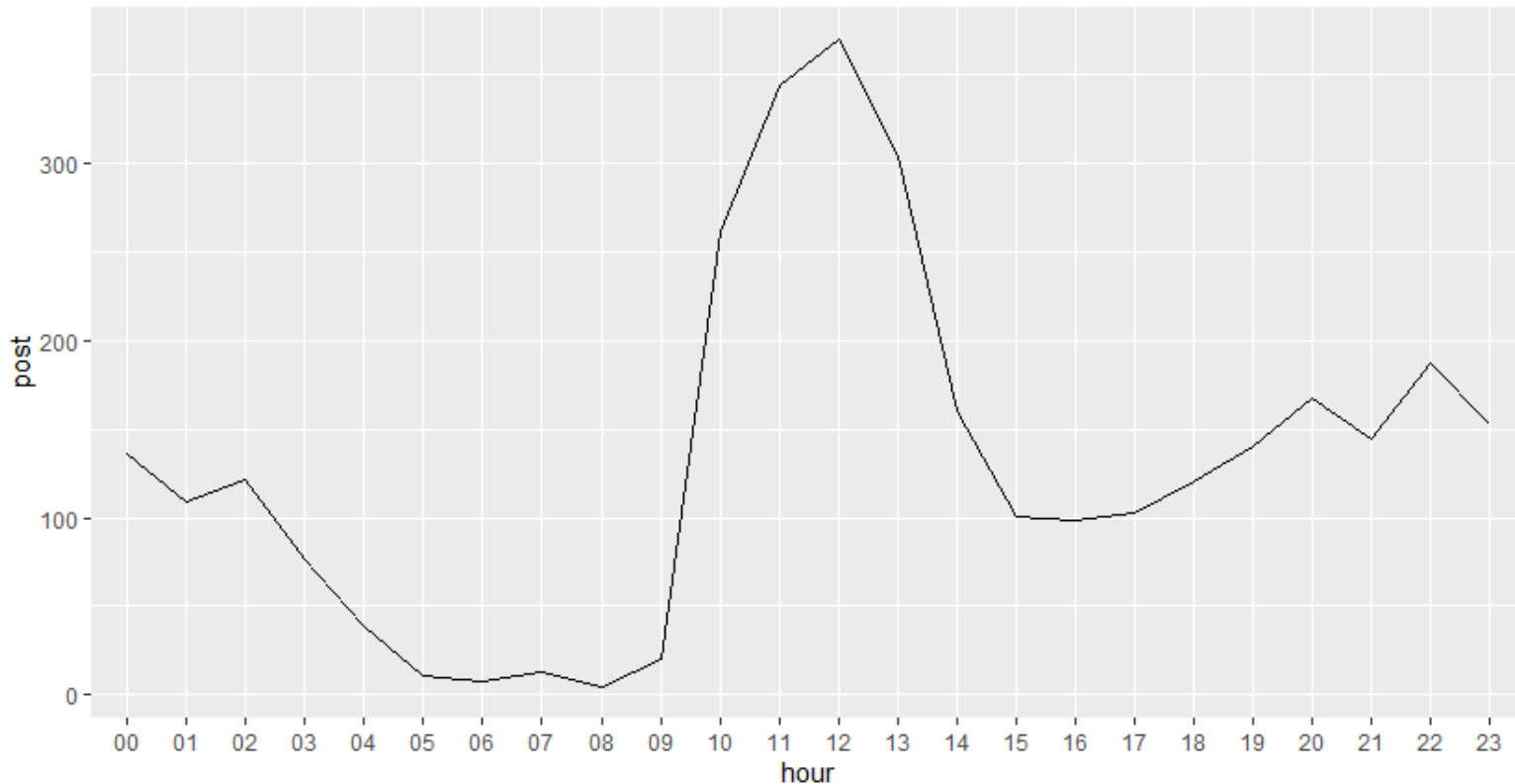
```
trump_hour <- trumptweets %>%  
  group_by(hour) %>%  
  summarise(post = n()) %>%  
  mutate(hour_rate = post / sum(post))
```

01

Time Series Analysis

Trump Tweets

```
ggplot(trump_hour, aes(x = hour, y = post, group = 1)) +  
  geom_line()
```



as.Date and as.POSIXct cannot produce monthly or quarterly data.

```
install.packages("zoo")  
library(zoo)
```

```
as.yearmon(as.Date("2015-04-07"))  
as.yearmon("2015-04-07") # This can produce correct result.  
I suggest you use the above codes.
```

```
as.yearqtr("2015-04-07")  
as.yearqtr(as.Date("2015-04-07"))
```


01

Time Series Analysis

Other Time Format

	Date	Arrivals
1	2014-02-05	100
2	2014-08-11	150
3	2014-11-23	200

Let's use `data_convert` to exercise how to convert date.

Date → month

Date → quarter

Month → the first and last date of month

Quarter → the first and last date of quarter

Create a plot to show the monthly number of Trump tweets.

	status_id	created_at	user_id	screen_name	text	source
1	880919871353217024	2017-06-30 22:43:58	25073877	realDonaldTrump	America will THINK BIG once again. We will inspire millions ...	Twitter Ads
2	880892632142143490	2017-06-30 20:55:44	25073877	realDonaldTrump	The era of strategic patience with the North Korea regime h...	Twitter for iPhone
3	880853246126022656	2017-06-30 18:19:13	25073877	realDonaldTrump	I am encouraged by President Moon's assurances that he wi...	Twitter Web Client
4	880771685460344832	2017-06-30 12:55:08	25073877	realDonaldTrump	Watched low rated @Morning_Joe for first time in long time...	Twitter for iPhone
5	880739852286918656	2017-06-30 10:48:38	25073877	realDonaldTrump	Crime and killings in Chicago have reached such epidemic p...	Twitter for iPhone
6	880737163247267840	2017-06-30 10:37:57	25073877	realDonaldTrump	If Republican Senators are unable to pass what they are wor...	Twitter for iPhone
7	880602793341591552	2017-06-30 01:44:01	25073877	realDonaldTrump	Just finished a very good meeting with the President of Sou...	Twitter for iPhone
8	880564232588107779	2017-06-29 23:10:47	25073877	realDonaldTrump	Our new American Energy Policy will unlock MILLIONS of jo...	Twitter Web Client
9	880563788705148928	2017-06-29 23:09:01	25073877	realDonaldTrump	When it comes to the future of America's energy needs, we ...	Twitter for iPhone
10	880559961855729664	2017-06-29 22:53:49	25073877	realDonaldTrump	Statement on House Passage of Kate's Law and No Sanctuar...	Twitter for iPhone
11	880553561008078848	2017-06-29 22:28:23	25073877	realDonaldTrump	Good news out of the House with the passing of 'No Sanctu...	Twitter for iPhone
12	880540637358981120	2017-06-29 21:37:02	25073877	realDonaldTrump	Good news, House just passed #KatesLaw. Hopefully Senate...	Twitter for iPhone
13	880534684479897600	2017-06-29 21:13:22	25073877	realDonaldTrump	"Mattis Says Trump's Warning Stopped Chemical Weapons ...	Twitter for iPhone
14	880508977590697984	2017-06-29 19:31:13	25073877	realDonaldTrump	Join me LIVE with @VP, @SecretaryPerry, @SecretaryZinke a...	Twitter for iPhone

```
trumptweets$ym <- as.yearmon(as.Date(trumptweets$created_at))
```

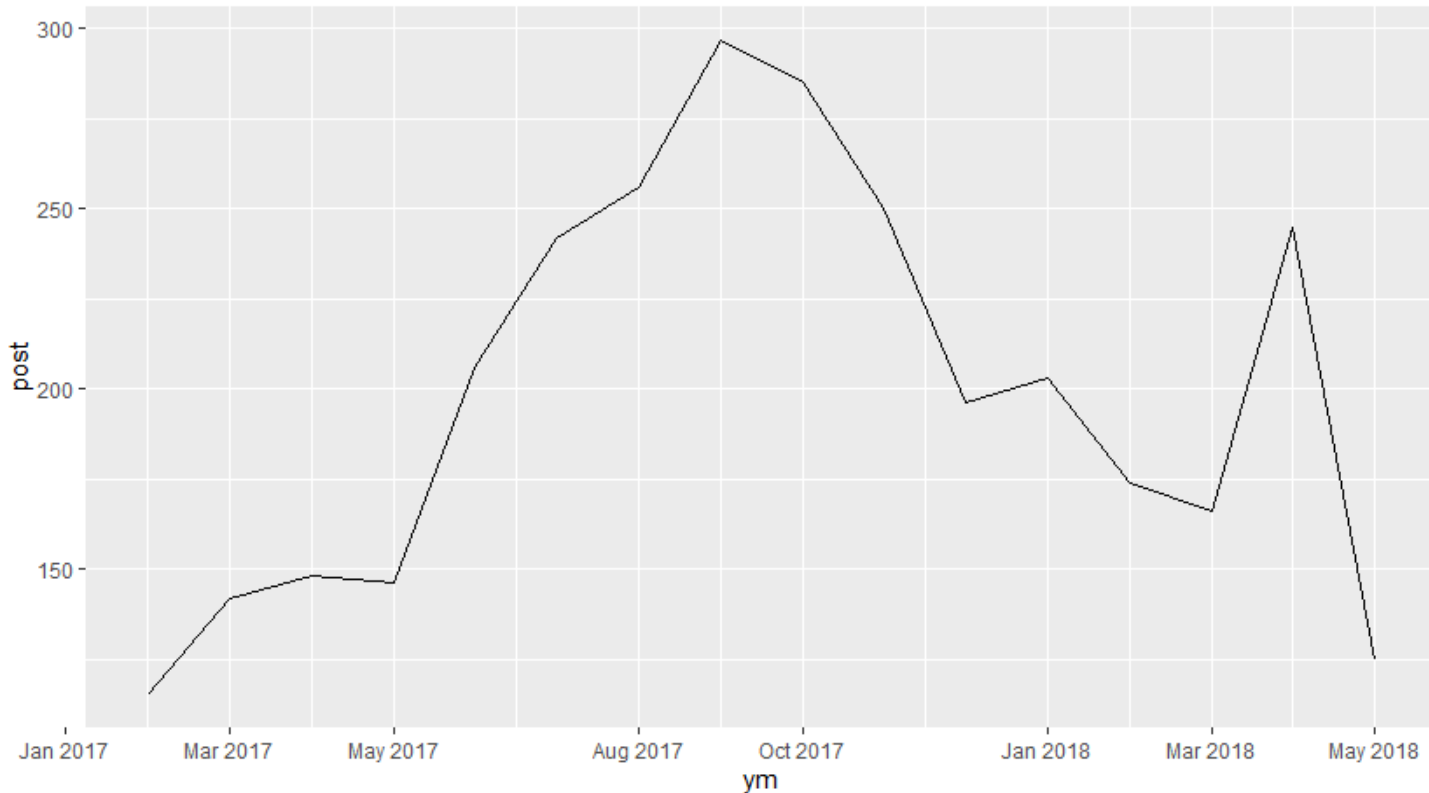
```
trump_ym <- trumptweets %>%  
  group_by(ym) %>%  
  summarise(post = n())
```

01

Time Series Analysis

Trump Tweets

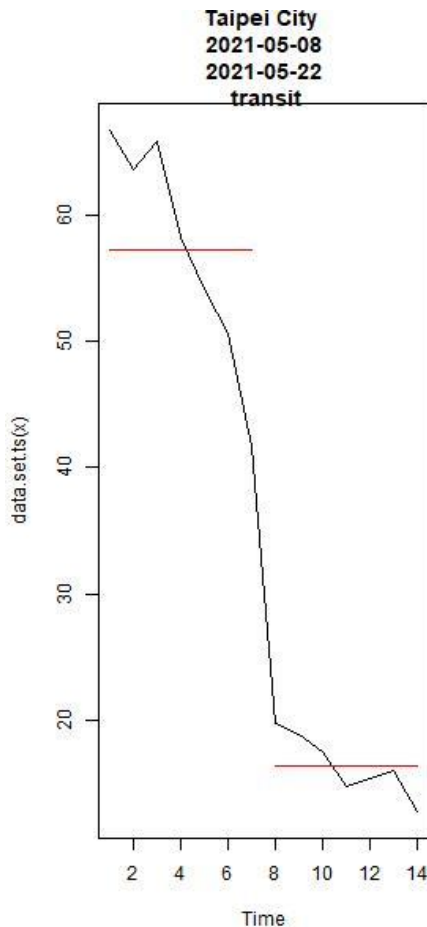
```
ggplot(trump_ym, aes(x = ym, y = post, group = 1)) +  
  geom_line()
```



Create a plot to show the quartly number of Trump tweets (Practice 2).

	status_id	created_at	user_id	screen_name	text	source
1	880919871353217024	2017-06-30 22:43:58	25073877	realDonaldTrump	America will THINK BIG once again. We will inspire millions ...	Twitter Ads
2	880892632142143490	2017-06-30 20:55:44	25073877	realDonaldTrump	The era of strategic patience with the North Korea regime h...	Twitter for iPhone
3	880853246126022656	2017-06-30 18:19:13	25073877	realDonaldTrump	I am encouraged by President Moon's assurances that he wi...	Twitter Web Client
4	880771685460344832	2017-06-30 12:55:08	25073877	realDonaldTrump	Watched low rated @Morning_Joe for first time in long time...	Twitter for iPhone
5	880739852286918656	2017-06-30 10:48:38	25073877	realDonaldTrump	Crime and killings in Chicago have reached such epidemic p...	Twitter for iPhone
6	880737163247267840	2017-06-30 10:37:57	25073877	realDonaldTrump	If Republican Senators are unable to pass what they are wor...	Twitter for iPhone
7	880602793341591552	2017-06-30 01:44:01	25073877	realDonaldTrump	Just finished a very good meeting with the President of Sou...	Twitter for iPhone
8	880564232588107779	2017-06-29 23:10:47	25073877	realDonaldTrump	Our new American Energy Policy will unlock MILLIONS of jo...	Twitter Web Client
9	880563788705148928	2017-06-29 23:09:01	25073877	realDonaldTrump	When it comes to the future of America's energy needs, we ...	Twitter for iPhone
10	880559961855729664	2017-06-29 22:53:49	25073877	realDonaldTrump	Statement on House Passage of Kate's Law and No Sanctuar...	Twitter for iPhone
11	880553561008078848	2017-06-29 22:28:23	25073877	realDonaldTrump	Good news out of the House with the passing of 'No Sanctu...	Twitter for iPhone
12	880540637358981120	2017-06-29 21:37:02	25073877	realDonaldTrump	Good news, House just passed #KatesLaw. Hopefully Senate...	Twitter for iPhone
13	880534684479897600	2017-06-29 21:13:22	25073877	realDonaldTrump	"Mattis Says Trump's Warning Stopped Chemical Weapons ...	Twitter for iPhone
14	880508977590697984	2017-06-29 19:31:13	25073877	realDonaldTrump	Join me LIVE with @VP, @SecretaryPerry, @SecretaryZinke a...	Twitter for iPhone

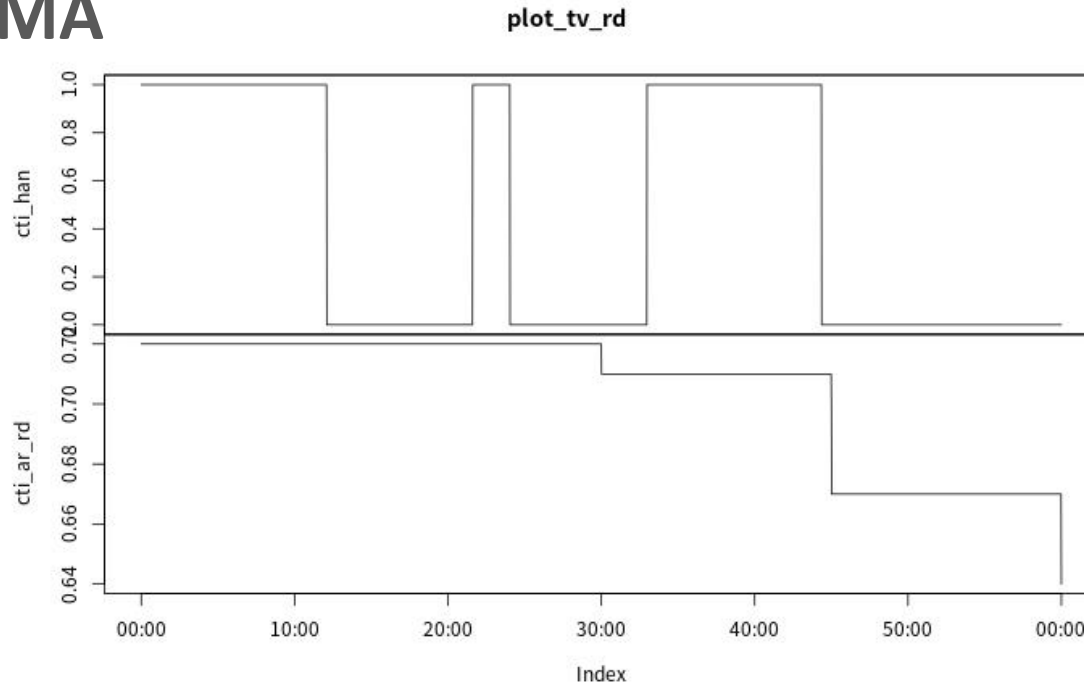
1. Changepoint analysis



`library(changepoint)`

To identify which points' changes are significant.

2. ARIMA



`library(forecast)`

x's changes affect y's changes?