

## ปฏิบัติการที่ 4

### การทำความสะอาดและจัดการข้อมูล

- วัตถุประสงค์:**
1. เพื่อให้นักศึกษาฝึกปฏิบัติทำความสะอาดและจัดการข้อมูลด้วยโปรแกรม R
  2. เพื่อให้นักศึกษาฝึกปฏิบัติจัดการตัวแปรรูปแบบวันที่ได้
  3. เพื่อให้นักศึกษาฝึกปฏิบัติผสมผสานข้อมูลจากหลาย ๆ ตารางได้

**Data file:** “unemployment.xlsx”, “child140\_practice.xlsx”, “thai\_road\_accident\_practice.csv”, “CBR64-Nara1\_practice.xlsx” และ “CBR64-Nara2\_practice.xlsx”

#### สรุปเนื้อหา:

การจัดเตรียมข้อมูล ได้แก่ การทำความสะอาดข้อมูล เป็นการขจัดข้อมูลที่ผิดพลาด ไม่สมบูรณ์ หรือซ้ำซ้อน การจัดรูปแบบข้อมูล เป็นการแปลงข้อมูลให้อยู่ในรูปแบบที่สอดคล้องกัน เช่น แปลงวันที่ให้อยู่ในรูปแบบเดียวกัน การจัดการข้อมูลที่ขาดหายไป เป็นการประมาณค่าหรือเติมข้อมูลที่ขาดหายไป การแปลงข้อมูลเป็นการแปลงข้อมูลให้เป็นรูปแบบที่เหมาะสมกับการวิเคราะห์ เช่น แปลงข้อมูลตัวอักษรเป็นตัวเลข การรวมข้อมูลเป็นการรวมข้อมูลจากหลายแหล่งเข้าด้วยกัน นอกจากการตรวจสอบข้อมูลข้างต้นแล้ว ยังมีการคำนวณค่าต่าง ๆ เพื่อเตรียมข้อมูลในการวิเคราะห์ข้อมูลต่อไป โดยทั่วไปที่เราจะตรวจสอบเป็นอันดับต้น ๆ คือ ข้อมูลซ้ำซ้อน ข้อมูลสูญหายซึ่งมักจะรายงานเป็นเปอร์เซ็นต์ ข้อมูลที่มีค่าไม่เป็นตามคำอธิบายข้อมูลซึ่งค่าเหล่านี้มักจะแทนค่าด้วย NA หรือหากผู้วิจัยสามารถตรวจสอบค่าจากแหล่งข้อมูลได้ก็สามารถแก้ไขข้อมูลที่บันทึกผิดพลาดให้ถูกต้องได้ การจัดการข้อมูลในกรณีนี้ที่ตัวแปรตามมีหลายตัวเพื่อให้ง่ายต่อการวิเคราะห์ข้อมูลและง่ายในการอธิบายผลการวิเคราะห์ข้อมูล

#### 1. ให้นักศึกษาฝึกจัดการโครงสร้างข้อมูลโดยใช้ data file “unemployment.xlsx”

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	
1	State	Area_name	y2007	y2008	y2009	y2010	y2011	y2012	y2013	y2014	y2015	y2016	y2017	y2018	
2	US	United States	4.6	5.8	9.3	9.6	9.0	8.1	7.4	6.2	5.3	4.9	4.4	3.9	
3	AL	Alabama	4.0	5.7	11.0	10.5	9.6	8.0	7.2	6.8	6.1	5.8	4.4	3.9	
4	AK	Alaska	6.3	6.7	7.7	7.9	7.6	7.1	7.0	6.9	6.5	6.9	7.0	6.6	
5	AZ	Arizona	3.9	6.2	9.9	10.4	9.5	8.3	7.7	6.8	6.1	5.4	4.9	4.8	
6	AR	Arkansas	5.3	5.5	7.8	8.2	8.3	7.6	7.2	6.0	5.0	4.0	3.7	3.7	
7	CA	California	5.4	7.3	11.2	12.2	11.7	10.4	8.9	7.5	6.2	5.5	4.8	4.2	
8	Co	Colorado	3.7	4.8	7.3	8.7	8.4	7.9	6.9	5.0	3.9	3.2	2.7	3.3	
9	DE	Delaware	3.4	4.9	8.3	8.4	7.5	7.2	6.7	5.7	4.9	4.5	4.5	3.8	
10	DC	District of Columbia	5.5	6.5	9.3	9.4	10.2	9.0	8.5	7.8	6.9	6.1	6.1	5.6	
11	FL	Florida	4.0	6.3	10.4	11.1	10.0	8.5	7.2	6.3	5.5	4.8	4.2	3.6	
12	GA	Georgia	4.5	6.2	9.9	10.5	10.2	9.2	8.2	7.1	6.0	5.4	4.7	3.9	
13	HI	Hawaii	2.8	4.3	7.2	6.9	6.8	6.0	4.9	4.4	3.6	3.0	2.4	2.4	
14	ID	Idaho	3.1	5.1	8.8	9.0	8.3	7.2	6.1	4.8	4.2	3.8	3.2	2.8	

a) ให้จัดข้อมูลจาก wide to long ดังตัวอย่างที่แสดงข้างล่างนี้

state	areaname	year	unemrate
US	United States	2007	4.6
US	United States	2008	5.8
US	United States	2009	9.3
US	United States	2010	9.6
US	United States	2011	9.0
...	...	...	...
AL	Alabama	2007	4.0
AL	Alabama	2008	5.7
AL	Alabama	2009	11.0
AL	Alabama	2010	10.5
AL	Alabama	2011	9.6
...	...	...	...

คำตอบ (แสดงคำสั่งและบางส่วนของผลลัพธ์)

```

Sys.setlocale("LC_ALL", locale="Thai")
# install.packages("readxl")
> library(readxl)
# install.packages("epiDisplay")
> library(epiDisplay)
# install.packages("reshape2")
> library(reshape2)

> setwd(".....")
> dt <- read_excel("unemployment.xlsx")
> dt <- as.data.frame(dt)
> str(dt)

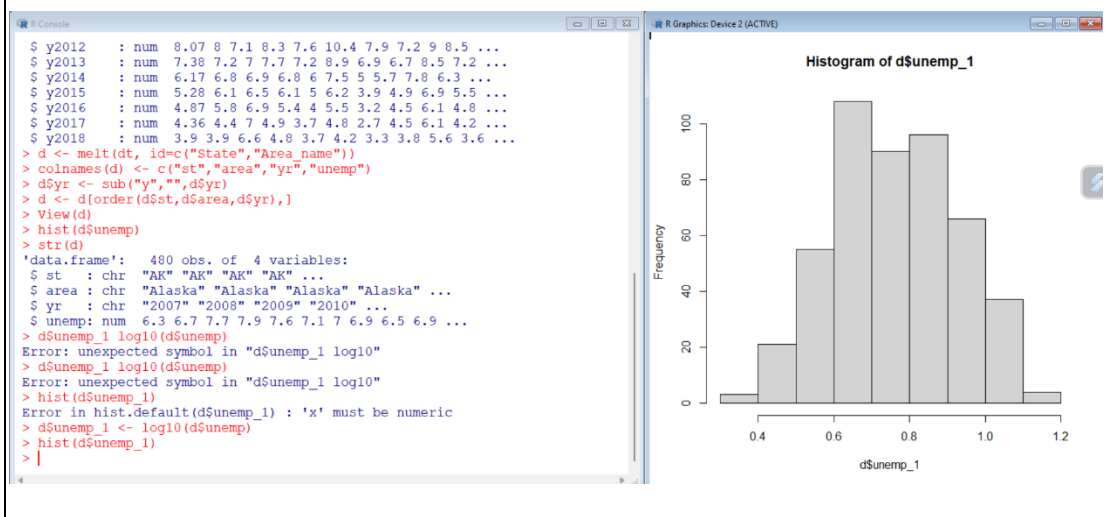
# Reshape
> d <- melt(dt, id=c("State", "Area_name"))
> colnames(d) <- c("st", "area", "yr", "unemp")
> d$yr <- sub("y", "", d$yr)
> d <- d[order(d$st, d$area, d$yr),]
> View(d)

```

- b) ให้ตรวจสอบการแจกแจงของข้อมูลอัตราการว่างงาน unemp หากข้อมูลไม่มีการแจกแจงแบบปกติให้ทำการแปลงข้อมูลให้อยู่ในรูปแบบการแจกแจงข้อมูลแบบปกติ

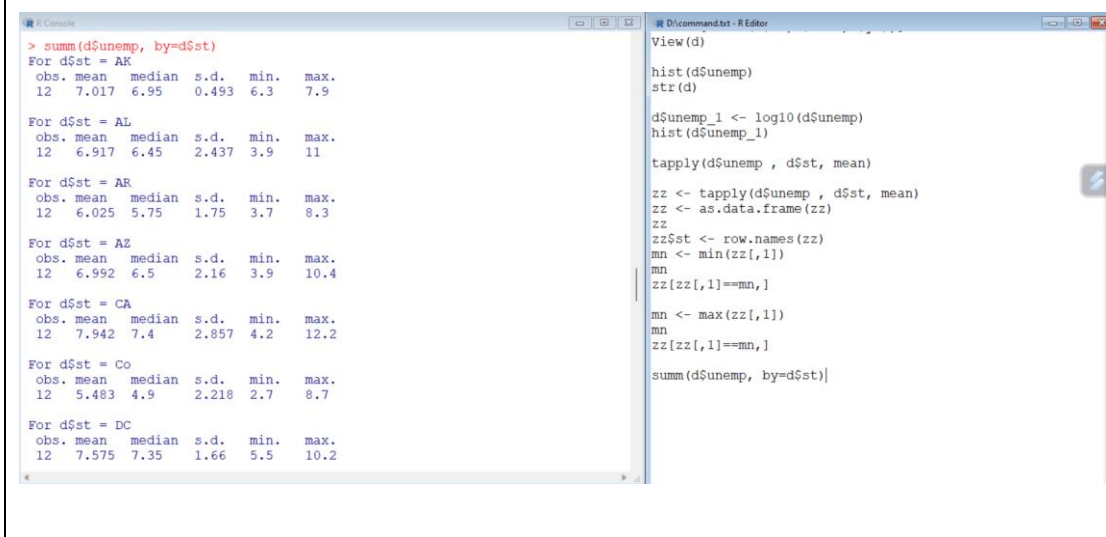
คำตอบ (แสดงคำสั่งและผลลัพธ์)

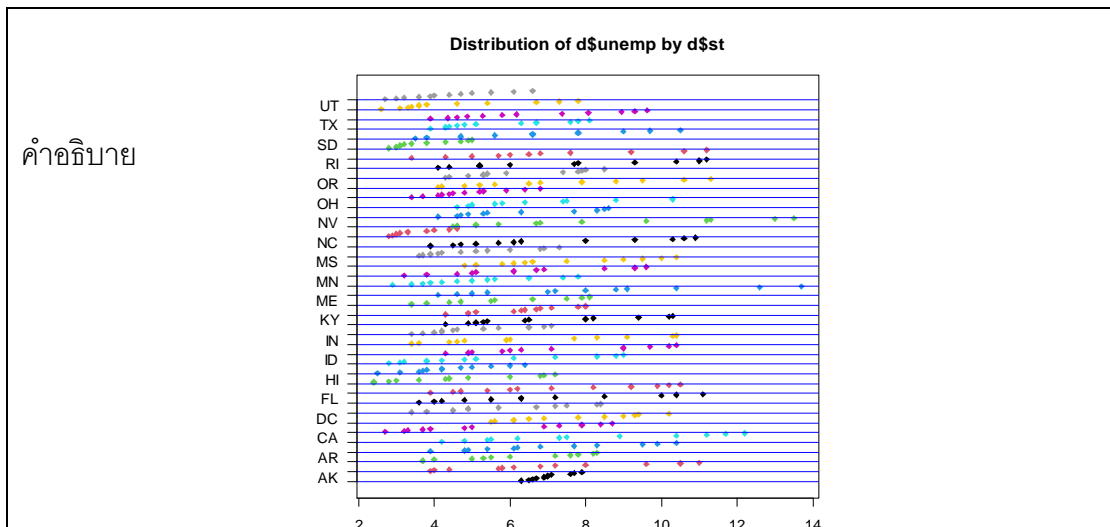
```
> hist(d$unemp)
```



- c) ให้ตรวจสอบข้อมูลแต่ละรัฐมีอัตราการว่างงานเป็นอย่างไรบ้าง โดยเฉลี่ยแล้วรัฐใดมีอัตราการว่างงานน้อยที่สุด และมากที่สุด

คำตอบ (แสดงคำสั่งและบางส่วนของผลลัพธ์)





- d) ให้บันทึกข้อมูลลงในไฟล์เดอร์ โดยตั้งชื่อไฟล์เป็น “unemploy2017.csv”

คำตอบ (แสดงคำสั่งและผลลัพธ์ที่ปรากฏในไฟล์เดอร์)

```
> write.csv(d, file="unemploy2017.csv", row.names=F)
```

2. ให้นักศึกษาฝึกจัดการข้อมูลตัวแปรวันที่โดยใช้ชุดข้อมูล “child140\_practice.xlsx”

	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	stu_sex	stu_birt	stu_stat	stu_weig	qtn_sibli	qtn_chil	stsmarn	brtdte	edulevr	occnme	sal	brtdte2	edulevr	occnme	sal2	qtn_sp2
2	1	2890	87	10	2	2	สมรส/มีคู่	2531101	ไม่ระบุ	รับจ้างทั่ว	4000	2531081	ไม่ระบุ	รับจ้างทั่ว	7000	3
3	1	3760	85	13	3	2	สมรส/มีคู่	2527031	ไม่ระบุ	ประมงแล	5000	2530070	มัธยมดอ	แม่บ้าน	0	4
4	1	2920	0	0	0		สมรส/มีคู่	2522082	มัธยมดอ	ค้าขาย/ป	10000	2527072	มัธยมดอ	ค้าขาย/ป	10000	2
5	2	3110	90	12	1	1	สมรส/มีคู่	2523092	ประถมศึกษา	รับจ้างทั่ว	6000	2531030	มัธยมดอ	รับจ้างทั่ว	6000	4
6	1	3310	90	12	4	4	สมรส/มีคู่	2522011	ไม่ระบุ	รับจ้างทั่ว	10000	2530010	ไม่ระบุ	แม่บ้าน	0	4
7	1	3132	0	0	0		สมรส/มีคู่	2521090	ประถมศึกษา	รับจ้างทั่ว	9000	2528020	ประถมศึกษา	แม่บ้าน	0	1
8	1	3590	0	0	0		สมรส/มีคู่	2525071	ไม่ระบุ	ไม่ระบุ	0	2527010	ไม่ระบุ	ลูกจ้างปร	7000	0
9	1	3010	0	0	0		สมรส/มีคู่	2529030	มัธยมดอ	รับจ้างทั่ว	6000	2530092	มัธยมดอ	รับจ้างทั่ว	10000	4
10	1	2920	90	12	1	1	สมรส/มีคู่	2540010	ประถมศึกษา	รับจ้างทั่ว	5000	2540052	มัธยมดอ	แม่บ้าน	0	4
11	1	3150	0	0	0	1	หม้าย	2533032	มัธยมดอ	รับจ้างทั่ว	9000	2535030	มัธยมดอ	แม่บ้าน	0	0
12	2	3145	89	12	1	1	สมรส/มีคู่	2534040	มัธยมดอ	รับจ้างทั่ว	10000	2534080	ไม่ระบุ	ไม่ระบุ	0	4
13	1	3010	0	0	0		สมรส/มีคู่	2531101	มัธยมดอ	รับจ้างทั่ว	7500	2532051	ปริญญาตรี	รับจ้างทั่ว	6000	0
14	1	3450	89	13	3	2	สมรส/มีคู่	2522063	ไม่ระบุ	พนักงาน	20000	2525112	ไม่ระบุ	ลูกจ้างปร	20000	0
15	2	3300	102	9.9	6	6	สมรส/มีคู่	2519012	มัธยมดอ	รับจ้างทั่ว	7000	2525011	มัธยมดอ	ค้าขาย/ป	6000	4
16	1	2780	89	13	3	2	สมรส/มีคู่	2535081	ไม่ระบุ	รับจ้างทั่ว	8000	2537060	ประถมศึกษา	แม่บ้าน	0	2
17	2	3600	95	12	2	2	สมรส/มีคู่	2536042	ไม่ระบุ	ลูกจ้างปร	7000	2537111	มัธยมดอ	ลูกจ้างปร	7000	4
18	2	3290	85	12	6	6	สมรส/มีคู่	2517061	มัธยมดอ	รับจ้างทั่ว	9000		ไม่ระบุ	แม่บ้าน	0	3
19	2	3130	92	12	1	1	สมรส/มีคู่	2531022	ไม่ระบุ	รับจ้างทั่ว	0	2526112	ไม่ระบุ	ไม่ระบุ	0	1
20	2	3280	83	10	1	1	สมรส/มีคู่	2538092	ไม่ระบุ	ค้าขาย/ป	9000	2540041	ปวส.	แม่บ้าน	0	3
21	1	2740	85	12	5	5	สมรส/มีคู่	2523092	ไม่ระบุ	ค้าขาย/ป	7000	2530060	ไม่ระบุ	แม่บ้าน	0	0
22	1	2960	89	13	2	2	สมรส/มีคู่	2516041	ไม่ระบุ	รับจ้างทั่ว	9000	2530051	ไม่ระบุ	แม่บ้าน	0	1
23	1	3080	90	12	2	1	สมรส/มีคู่		ไม่ระบุ	รับจ้างทั่ว	6000	2536122	ไม่ระบุ	แม่บ้าน	0	3
24	2	3030	96	12	2	2	สมรส/มีคู่	2530030	ไม่ระบุ	ลูกจ้างปร	20000	2528071	ไม่ระบุ	ลูกจ้างปร	20000	4
25	2	3450	105	13	4	4	สมรส/มีคู่	2517071	ไม่ระบุ	รับจ้างทั่ว	9000	2523032	ไม่ระบุ	ค้าขาย/ป	6000	1
26	2	3230	85	12	3	3	สมรส/มีคู่	2515012	ปริญญาตรี	ข้าราชการ	28000	2530051	ปริญญาตรี	รับจ้างทั่ว	6000	0

# กำหนดไฟล์เดอร์ที่มีข้อมูล “child140\_practice.xlsx”

```

> setwd(".....")

# เรียกใช้ library excel

> library(readxl)

# อ่าน data file "child140_practice.xlsx")

> dt <- read_excel("child140_practice.xlsx")

> dt <- data.frame(dt)

> str(dt)

'data.frame': 140 obs. of 32 variables:
 $ stu_prenme      : chr  "ด.ช." "ด.ช." "ด.ช." "ด.ญ." ...
 $ stu_brtde       : chr  "25610303" "25610403" "25610314"
 $ stu_sex         : chr  "1" "1" "1" "2" ...
 $ stu_birth_weight : chr  "2890" "3760" "2920" "3110" ...
 $ stu_stature     : chr  "87" "85" "0" "90" ...
 $ stu_weight      : chr  "10" "13" "0" "12" ...
 $ qtn_siblings_faandmom: chr  "2" "3" "0" "1" ...
 $ qtn_childno     : chr  "2" "2" NA "1" ...
 $ stsmarname      : chr  "สมรส/มีคู่/อยู่ด้วยกัน" "สมรส/มีคู่/อยู่ด้วยกัน" "สมรส/มีคู่/
 $ brtde           : chr  "25311010" "25270310" "25220826"
 $ edulevnmel      : chr  "ไม่ระบุ" "ไม่ระบุ" "มัธยมศึกษาตอนปลาย" "ประถมศึกษา" ...
 $ occnme          : chr  "รับจ้างทั่วไป/กรรมกร" "ประมงและการเดินเรือ" "ค้าขาย/ปร
 $ sal             : num  4000 5000 10000 6000 10000 9000 (
 $ brtde2          : chr  "25310810" "25300703" "25270727"

```

- a) ให้คำนวณอายุของกลุ่มตัวอย่างแต่ละคน (ตัวแปร stu\_brtde) เก็บไว้ในตัวแปรใหม่ชื่อ stuage โดยแสดงในรูปแบบตาราง

คำตอบ (แสดงคำสั่งและผลลัพธ์)

```

> dt$stubrtdte <- as.Date(dt$stu_brtde, "%Y%m%d")

> today <- as.Date("25670707", "%Y%m%d")

> dt$stuage <- as.numeric(today - dt$stubrtdte)/365.25

> str(dt)

 $ qtn_sp2_7      : chr  "2" "3" "0" "1" ...
 $ qtn_sp2_8      : chr  "1" "4" "0" "1" ...
 $ qtn_sp2_9      : chr  "2" "4" "1" "4" ...
 $ qtn_sp2_10     : chr  "2" "4" "0" "4" ...
 $ qtn_sp2_11     : chr  "2" "4" "1" "2" ...
 $ qtn_sp2_12     : chr  "2" "3" "2" "1" ...
 $ qtn_sp2_13     : chr  "2" "3" "1" "0" ...
 $ stubrtdte      : Date, format: "2561-03-03"
 $ stuage         : num  6.34 6.26 6.31 7.11 6.

> summary(dt$stuage)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 5.563   5.884   6.142   6.241   6.611   7.507

```

- ```

> summary(dt$fage)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
5.563   5.884   6.142   6.241   6.611   7.507
> dt$fbrtdate <- as.Date(dt$fbrtdate2, "%Y%m%d")
Error in `<-data.frame`(`*tmp*`, fbrtdate, value = numeric(0)) :
  replacement has 0 rows, data has 140
> dt$fage <- as.numeric(today - dt$fbrtdate)/365.25
> summary(dt$fage)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
5.563   5.884   6.142   6.241   6.611   7.507
> dt$fbrtdate <- as.Date(dt$brrtdate2, "%Y%m%d")
> dt$fage <- as.numeric(today - dt$fbrtdate)/365.25
> summary(dt$fage)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max. NA's
-517.69  30.86   35.73   31.42   40.11   50.35   11
> dt$fbrtdate <- as.Date(dt$brrtdate, "%Y%m%d")
> dt$fage <- as.numeric(today - dt$fbrtdate)/365.25
> summary(dt$fage)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
5.563   5.884   6.142   6.241   6.611   7.507
> dt$monbrrtdate <- as.Date(dt$brrtdate2, "%Y%m%d")
> dt$smomage <- as.numeric(today - dt$monbrrtdate)/365.25
> summary(dt$smomage)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max. NA's
-517.69  30.86   35.73   31.42   40.11   50.35   11
> dt[which(dt$smomage<0),]
  stu_prenme stu_brrtdate stu_sex stu_birth_weight stu_stature stu_weight
134   a.e. 25610330      1      3660       100       18

```

- ```

คำตอบ (แสดงคำสั่งและผลลัพธ์)

Error in `<-data.frame`(`*tmp*`, fabrtdate, value = numeric(0)) :
  replacement has 0 rows, data has 140
> dt$faage <- as.numeric(today - dt$fabrtdate)/365.25
> summary(dt$faage)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
5.563   5.884   6.142   6.241  6.611   7.507
> dt$fabrtdate <- as.Date(dt$brtdate2, "%Y%m%d")
> dt$faage <- as.numeric(today - dt$fabrtdate)/365.25
> summary(dt$faage)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
-517.69  30.86   35.73  31.42  40.11   50.35    11
> dt$fabrtdate <- as.Date(dt$stu_brtdate, "%Y%m%d")
> dt$faage <- as.numeric(today - dt$fabrtdate)/365.25
> summary(dt$faage)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
5.563   5.884   6.142   6.241  6.611   7.507
> dt$smombrtdate <- as.Date(dt$brtdate2, "%Y%m%d")
> dt$smomage <- as.numeric(today - dt$smombrtdate)/365.25
> summary(dt$smomage)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
-517.69  30.86   35.73  31.42  40.11   50.35    11
> dt[which(dt$smomage<0),]
  stu_prenme stu_brtdate stu_sex stu_birth_weight stu_stature stu_weight
134  อ.ช. 25610330   1      3660    100   18
    qtn_siblings faandmom qtn_childno      stsmarnme   brtdate
134      1      1  ศาสตราจารย์ชายคน 25420607
    edulevnm1 occnm1 sal brtdate2 edulevnm2 occnm2 sal2 qtn_sp2_1
134  โสภณ โสภณ 6000 30850321 มีบุตรสองคน โสภณ 0   0

setwd("D:\\mcs")
dt <- read_excel("child140_practice (2).xlsx")
dt <- as.data.frame(dt)
str(dt)

dt$stubrtdte <- as.Date(dt$stu_brtdate, "%Y%m%d")
today <- as.Date("25670707", "%Y%m%d")
dt$stuage <- as.numeric(today - dt$stubrtdte)/365.25
str(dt)

summary(dt$stuage)

dt$fabrtdate <- as.Date(dt$stu_brtdate, "%Y%m%d")
dt$faage <- as.numeric(today - dt$fabrtdate)/365.25
summary(dt$faage)

dt$smombrtdate <- as.Date(dt$brtdate2, "%Y%m%d")
dt$smomage <- as.numeric(today - dt$smombrtdate)/365.25
summary(dt$smomage)

dt[which(dt$smomage<0),]
dt$smomage <- ifelse (dt$smomage < 0 ,NA,dt$smomage)
summary(dt$smomage)

```

```

      stuage  fabrtdate  faage  mombrdate  momage
134 6.269678 2561-03-30 6.269678 3085-03-21 -517.6947
> dt$momage <- ifelse (dt$momage < 0 , NA, dt$momage)
> summary(dt$momage)
Error in summary(dt$momage) : could not find function "summary"
> dt$momage <- ifelse (dt$momage < 0 , NA, dt$momage)
> summary(dt$momage)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
23.02  31.54   35.82  35.71  40.11   50.35    12
> dt[which(dt$momage<0),]
   [1] stu_prenme  stu_brtdate  stu_sex
   [4] stu_birth_weight  stu_stature  stu_weight
   [7] qtn_siblings_faandmom  qtn_childno  stsmarname
  [10] brtdate  edulevnmel  occnme
  [13] sal  brtdate2  edulevnmel2
  [16] occnme2  sal2  qtn_sp2_1
  [19] qtn_sp2_2  qtn_sp2_3  qtn_sp2_4
  [22] qtn_sp2_5  qtn_sp2_6  qtn_sp2_7
  [25] qtn_sp2_8  qtn_sp2_9  qtn_sp2_10
  [28] qtn_sp2_11  qtn_sp2_12  qtn_sp2_13
  [31] stubrtdate  stuage  fabrtdate
  [34] faage  mombrdate  momage
<0 rows> (or 0-length row.names)
> dt$momage <- ifelse (dt$momage < 0 , NA, dt$momage)
> summary(dt$momage)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
23.02  31.54   35.82  35.71  40.11   50.35    12
>

```

d) ให้จัดกลุ่มอายุบิดา ตามเงื่อนไขที่แสดงข้างล่างนี้ เก็บไว้ในตัวแปรใหม่ชื่อ faagegrp โดยแสดงในรูปแบบตารางหรือกราฟ

- 1: <=30
- 2: 31-40
- 3: >40
- 4 Undefined (กรณีไม่ระบุ/Missing value)

คำตอบ (แสดงคำสั่งและผลลัพธ์)

```

> dt$faagegrp <- ifelse(is.na(dt$faage), 4,
+ ifelse(dt$faage<=30, 1,
+ ifelse(dt$faage<=40, 2, 3)))
> tab1(dt$faagegrp)
dt$faagegrp :
      Frequency Percent Cum. percent
1          16      11.4         11.4
2          53      37.9         49.3
3          55      39.3         88.6
4          16      11.4        100.0
Total       140     100.0        100.0
> dt$faagegrp <- factor(dt$faagegrp)
> levels(dt$faagegrp) <- c("1:<=30", "2:31-40", "3:>40", "4:Undefined")
> tab1(dt$faagegrp)
dt$faagegrp :
      Frequency Percent Cum. percent
1:<=30          16      11.4         11.4
2:31-40         53      37.9         49.3
3:>40          55      39.3         88.6
4:Undefined     16      11.4        100.0
Total       140     100.0        100.0

```

```

dt$momage <- ifelse (dt$momage < 0 , NA, dt$momage)
summary(dt$momage)

dt$faagegrp <- ifelse(is.na(dt$faage), 4,
                      ifelse(dt$faage<=30, 1,
                              ifelse(dt$faage<=40, 2, 3)))
tab1(dt$faagegrp)

dt$faagegrp <- factor(dt$faagegrp)
levels(dt$faagegrp) <- c("1:<=30", "2:31-40", "3:>40", "4:Undefined")
tab1(dt$faagegrp)

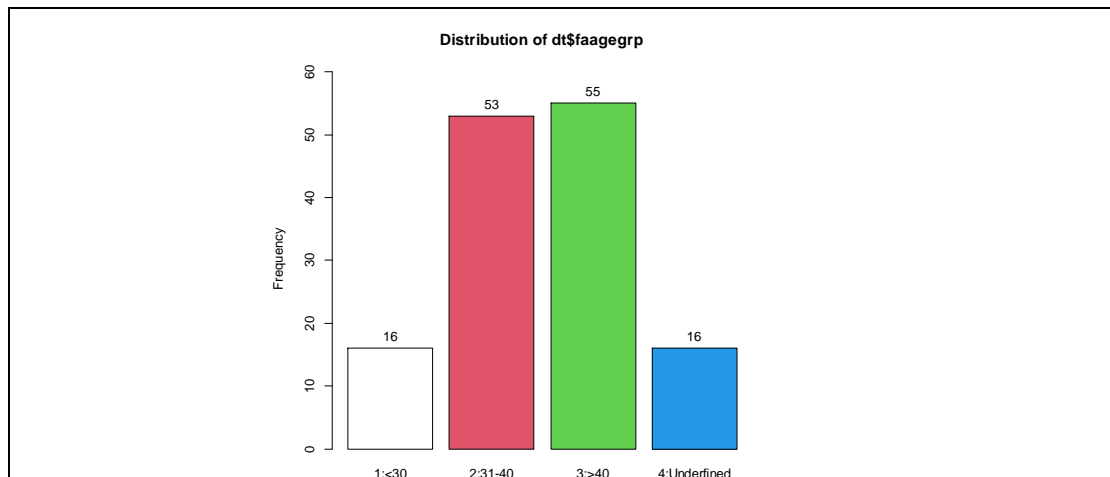
dt$momagegrp <- dt$momage
dt$momagegrp <- ifelse (dt$momage<=30, 1,
                       ifelse (dt$momage<=40, 2, 3))

tab1(dt$momagegrp)

```

คำอธิบายตาราง





e) ให้จัดกลุ่มอายุมารดา ตามเงื่อนไขที่แสดงข้างล่างนี้ เก็บไว้ในตัวแปรใหม่ชื่อ momagegrp โดยแสดงผลในรูปแบบตารางหรือกราฟ

1: <=30

2: 31-40

3: >40

4 Undefined (กรณีไม่ระบุ/Missing value)

คำตอบ (แสดงคำสั่งและผลลัพธ์)

```

Frequency % (NA+) % (NA-)
1          24      17.1    18.8
2          69      49.3    53.9
3          35      25.0    27.3
<NA>       12       8.6     0.0
Total     140     100.0   100.0
> dt$faagegrp <- factor(dt$faagegrp)
> levels(dt$faagegrp) <- c("1:<=30", "2:31-40", "3:>40", "4:Undefined")
> tab1(dt$faagegrp)
dt$faagegrp :
Frequency Percent Cum. percent
1:<=30         16      11.4      11.4
2:31-40       53      37.9      49.3
3:>40         55      39.3      88.6
4:Undefined   16      11.4     100.0
Total        140     100.0
> dt$momagegrp <- dt$momage
> dt$momagegrp <- ifelse(dt$momage <= 30, 1,
+ ifelse(dt$momage > 30, 2, 3))
> tab1(dt$momagegrp)
dt$momagegrp :
Frequency % (NA+) % (NA-)
1          24      17.1    18.8
2          69      49.3    53.9
3          35      25.0    27.3
<NA>       12       8.6     0.0
Total     140     100.0   100.0
>

```

```

dt$mombtrdte <- as.Date(dt$btrdte2, "%Y%m%d")
dt$momage <- as.numeric(today - dt$mombtrdte)/365.25
summary(dt$momage)

dt[which(dt$momage < 0),]
dt$momage <- ifelse(dt$momage < 0, NA, dt$momage)
summary(dt$momage)

dt$faagegrp <- ifelse(is.na(dt$faage), 4,
  ifelse(dt$faage <= 30, 1,
    ifelse(dt$faage > 30, 2, 3)))
tab1(dt$faagegrp)

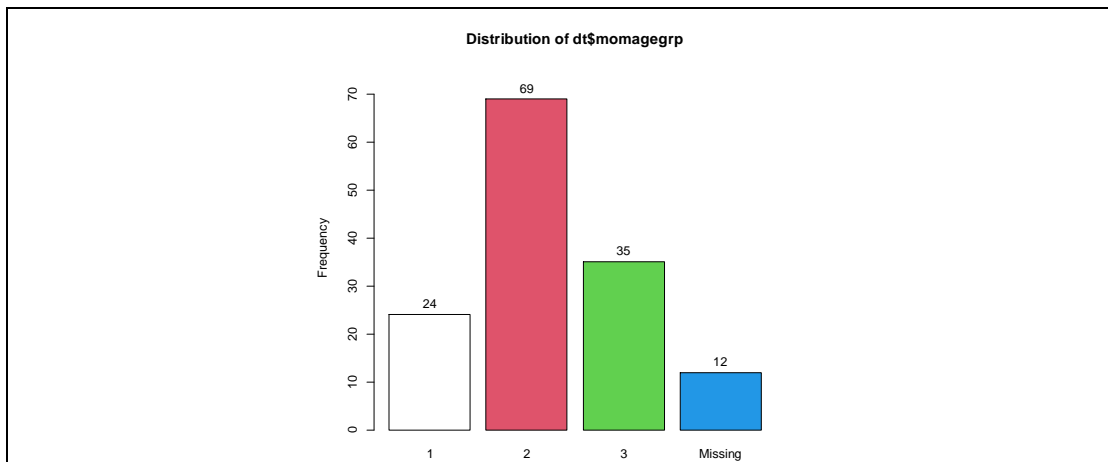
dt$faagegrp <- factor(dt$faagegrp)
levels(dt$faagegrp) <- c("1:<=30", "2:31-40", "3:>40", "4:Undefined")
tab1(dt$faagegrp)

dt$momagegrp <- dt$momage
dt$momagegrp <- ifelse(dt$momage <= 30, 1,
  ifelse(dt$momage > 30, 2, 3))
tab1(dt$momagegrp)

```

คำอธิบายตาราง





f) แต่ละกลุ่มอายุของบิดามีเงินเดือนเฉลี่ยเท่าไร? แสดงผลลัพธ์ในรูปแบบตาราง

คำตอบ (แสดงคำสั่งและผลลัพธ์)

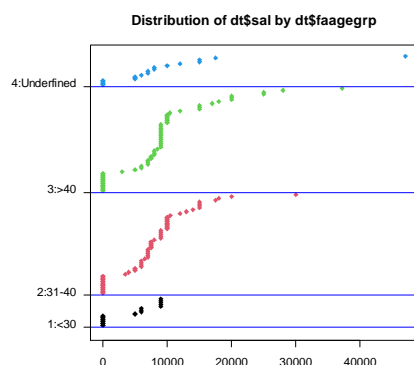
```
> tapply(dt$sal, dt$faagegrp, mean, na.rm=T)
1:<30 2:31-40 3:>40 4:Underfined
4533.333 8132.075 9692.545 10156.250
> sum(dt$faagegrp)
For dt$faagegrp = 1:<30
obs. mean median s.d. min. max.
15 4533.333 6000 4050.867 0 9000
For dt$faagegrp = 2:31-40
obs. mean median s.d. min. max.
53 8132.075 7500 5961.536 0 30000
For dt$faagegrp = 3:>40
obs. mean median s.d. min. max.
55 9692.545 9000 7665.467 0 37190
For dt$faagegrp = 4:Underfined
obs. mean median s.d. min. max.
16 10156.25 7500 11167.988 0 47000
> |
```

```
dt$momagegrp <- ifelse(dt$momage <= 30, 1,
  ifelse(dt$momage <= 40, 2, 3))
tbl1(dt$momagegrp)

# 2variable categorical vs continuous=====
tapply(dt$sal, dt$faagegrp, mean, na.rm=T)
sum(dt$sal, by=dt$faagegrp)

# 2variable categorical vs continuous=====
table(dt$occnme, dt$faagegrp)
tabpct(dt$occnme, dt$faagegrp, percent="row")
```

คำอธิบายตาราง



g) แต่ละกลุ่มอายุของมารดามีเงินเดือนเฉลี่ยเท่าไร? แสดงผลลัพธ์ในรูปแบบตาราง

คำตอบ (แสดงคำสั่งและผลลัพธ์)

```

> mom_sal <- tapply(dt$sal, dt$momagegrp, mean, na.rm = TRUE)
> mom_sal
      1:<30      2:31-40      3:>40 4:Underfined
6270.833  8933.824  9359.714      NA
> summa(dt$sal, by=dt$momagegrp)
Error in par(mai = c(0.95625, left.offet, 0.76875, 0.39375)) :
invalid value specified for graphical parameter "mai"
> summa(dt$sal, by=dt$mom_sal)
obs. mean median s.d. min. max.
139 8594.173 8000 7365.088 0 47000
> summa(dt$mom_sal, by=dt$momagegrp)
Error in order(c(2L, 2L, 2L, 2L, 2L, 3L, 2L, 1L, 2L, 2L, 3L, 3L, :
argument lengths differ
> summa(dt$mom_sal, by=dt$sal)
Error in order(by, as.numeric(x)) : argument lengths differ
> summa(dt$sal, by=dt$mom_sal)
obs. mean median s.d. min. max.
139 8594.173 8000 7365.088 0 47000

```

```

mom_sal <- tapply(dt$sal, dt$momagegrp, mean, na.rm = TRUE)
mom_sal
summa(dt$sal, by=dt$mom_sal)
table(dt$edu, dt$momagegrp)
tabpct(dt$edu, dt$momagegrp, percent="row")

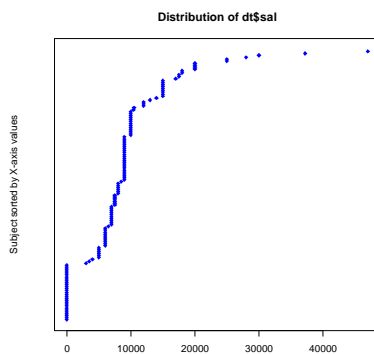
#=====
nr_1 <- read_excel("CBR64-Nara1_practice.xlsx")
nr_1 <- as.data.frame(nr_1)
str(nr_1)

nr_2 <- read_excel("CBR64-Nara2_practice.xlsx")
nr_2 <- as.data.frame(nr_2)
str(nr_2)

nr_3 <- rbind(nr_1, nr_2)

```

### คำอธิบายตาราง



h) แต่ละกลุ่มอายุของบิดาประกอบอาชีพอะไรบ้าง แสดงผลลัพธ์ในรูปแบบตาราง

### คำตอบ (แสดงคำสั่งและผลลัพธ์)

```

> tabpct(dt$occnme, dt$faagegrp, percent="row")

Row percent
dt$occnme dt$faagegrp
1:<30 2:31-40 3:>40 4:Underfined Total
ข้าราชการ 0 0 2 0 2 (0) (100)
ข้าราชการเกษียณอายุ 0 0 0 1 1 (0) (100)
ค้าขาย/ประกอบธุรกิจส่วนตัว 2 2 5 3 10 (16.7) (41.7) (25) (100)
พนักงานบริษัทเอกชน 1 1 3 0 5 (20) (60) (0) (100)
พนักงานของรัฐ 0 1 1 0 2 (50) (50) (0) (100)
พนักงานรัฐวิสาหกิจ 0 1 1 0 2 (50) (50) (0) (100)
ไม่ระบุ 8 22 21 5 56 (14.3) (39.3) (37.5) (8.9) (100)
รับจ้างทั่วไป/กรรมกร 5 19 20 6 50 (10) (38) (40) (12) (100)
ลูกจ้างบริษัท/พนักงานเอกชน 0 1 0 0 1 (100) (0) (0) (100)
ลูกจ้างประจำ 0 5 0 0 5 (100) (0) (0) (100)
อาชีพอิสระ/ฟรีแลนซ์ (อา.) 0 1 1 1 3 (33.3) (33.3) (33.3) (100)

```

```

summary(dt$momage)

dt$faagegrp <- ifelse(is.na(dt$faage), 4,
  ifelse(dt$faage <= 30, 1,
    ifelse(dt$faage <= 40, 2, 3)))
tabl(dt$faagegrp)

dt$faagegrp <- factor(dt$faagegrp)
levels(dt$faagegrp) <- c("1:<30", "2:31-40", "3:>40", "4:Underfined")
tabl(dt$faagegrp)

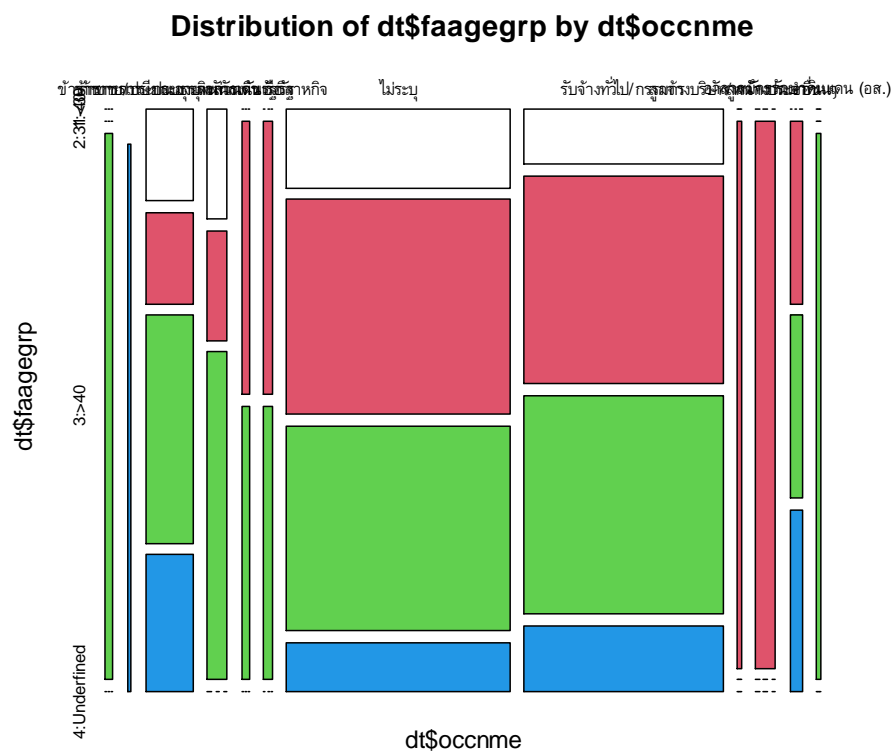
dt$momagegrp <- dt$momage
dt$momagegrp <- ifelse(dt$momage <= 30, 1,
  ifelse(dt$momage <= 40, 2, 3))
tabl(dt$momagegrp)

# 2variable categorical vs continuous=====
tapply(dt$sal, dt$faagegrp, mean, na.rm=T)
summa(dt$sal, by=dt$faagegrp)

# 2variable categorical vs continuous=====
table(dt$occnme, dt$faagegrp)
tabpct(dt$occnme, dt$faagegrp, percent="row")

```

คำอธิบายตาราง



- i) แต่ละกลุ่มอายุของมารดาจบการศึกษาระดับใดบ้าง แสดงผลลัพธ์ในรูปแบบตาราง

คำตอบ (แสดงคำสั่งและผลลัพธ์)

```

> table(dt$edulevnmel , dt$momagegrp)
      1:<30 2:31-40 3:>40 4:Undefined
ประถมศึกษา  7  9  7  0
ปริญญาตรี  1 11  6  0
ปริญญาโท    0  0  1  0
ปวส.        1  0  0  0
ปวช.        1  7  0  0
มัธยมศึกษา  2 11  3  0
มัธยมศึกษา  4 13  6  0
ไม่จบระดับประถมศึกษา 1  0  0  0
ไม่ระบุ     7 18 12  0
> tabpct(dt$edulevnmel , dt$momagegrp, percent="row")
Row percent
dt$edulevnmel      dt$momagegrp
      1:<30 2:31-40 3:>40 4:Undefined Total
ประถมศึกษา  7  9  7  0 23
      (30.4) (39.1) (30.4)      (0) (100)
ปริญญาตรี  1 11  6  0 18
      (5.6) (61.1) (33.3)      (0) (100)
ปริญญาโท    0  0  1  0  1
      (0)      (0) (100)      (0) (100)
ปวส.        1  0  0  0  1
      (100)      (0) (0)      (0) (100)
ปวช.        1  7  0  0  8
      (12.5) (87.5) (0)      (0) (100)
มัธยมศึกษา  2 11  3  0 16
      (12.5) (68.8) (18.8)      (0) (100)
มัธยมศึกษา  4 13  6  0 23
      (17.4) (56.5) (26.1)      (0) (100)
ไม่จบระดับประถมศึกษา 1  0  0  0  1
      (100)      (0) (0)      (0) (100)

```

```

str(dt)
table(dt$edulevnmel , dt$momagegrp)
tabpct(dt$edulevnmel , dt$momagegrp, percent="row")

#=====
nr_1 <- read_excel("CBR64-Nara1_practice.xlsx")
nr_1 <- as.data.frame(nr_1)
str(nr_1)

nr_2 <- read_excel("CBR64-Nara2_practice.xlsx")
nr_2 <- as.data.frame(nr_2)
str(nr_2)

nr_3 <- rbind(nr_1,nr_2)

des(nr_1)
des(nr_2)

nr_11 <- nr_1[,c("tit","sex","stt",
"brtdte","tmb","prv","edu","occ","mem",
"inc53m","inc55y")]

nr_22 <- nr_2[,c("tit","sex","stt",
"brtdte","tmb","prv","edu","occ","mem",
"inc53m","inc55y")]

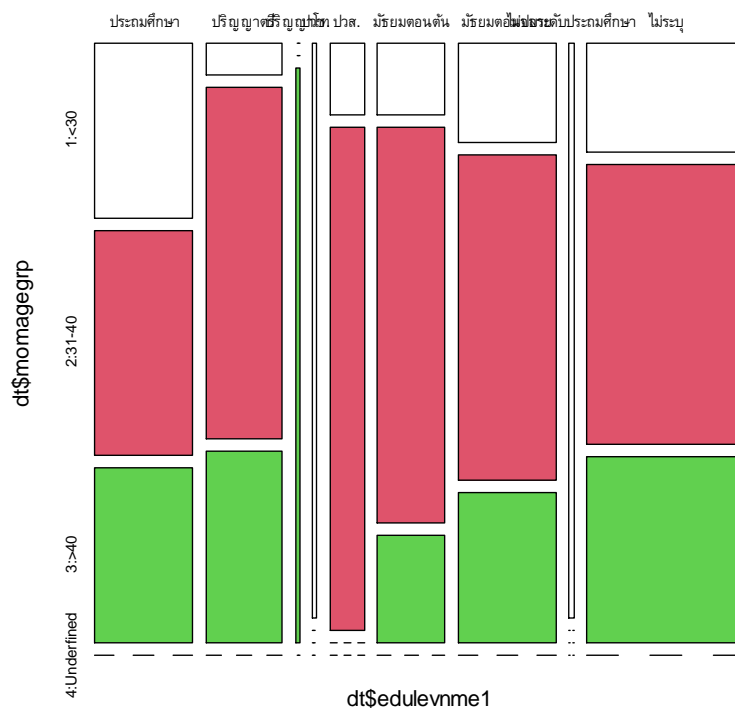
des(nr_11)
des(nr_22)

nr_3 <- rbind(nr_11,nr_22)
des(nr_3)
str(nr_3)

```

คำอธิบายตาราง

**Distribution of dt\$momagegrp by dt\$edulevnmel**



## 3. ให้นักศึกษาฝึกสำรวจข้อมูลโดยใช้ชุดข้อมูล "thai\_road\_accident\_practice.csv"

	A	B	C	D	E	F	G	H	I	J
1	acc_code	incident_datetime	report_datetime	province_en	agency	vehicle_type	presumed_cause	accident_type	number_of_vehicles_involved	number_of_fatalities
2	571905	1/1/2019 0:00	2/1/2019 6:11	Loburi	department of rural roads	motorcycle	driving under the influence of alcohol	other	1	0
3	3790870	1/1/2019 0:03	20/2/2020 13:48	Ubon Ratchathani	department of highways	private/passenger car	speeding	rollover/fallen on straight road	1	0
4	599075	1/1/2019 0:05	1/1/2019 10:35	Prachuap Khiri Khan	department of rural roads	motorcycle	speeding	head-on collision	2	1
5	571924	1/1/2019 0:20	2/1/2019 5:12	Chiang Mai	department of rural roads	motorcycle	driving under the influence of alcohol	other	1	0
6	599523	1/1/2019 0:25	4/1/2019 9:42	Nakhon Si Thammarat	department of rural roads	private/passenger car	cutting in	rollover/fallen on straight road	1	0
7	571982	1/1/2019 0:30	7/1/2019 12:46	Mae Hong Son	department of rural roads	motorcycle	speeding	other	1	0
8	612782	1/1/2019 0:30	25/10/2019 14:25	Chumphon	department of rural roads	4-wheel passenger vehicle	failure to yield	collision at intersection	2	0
9	599235	1/1/2019 0:35	2/1/2019 16:23	Sing Buri	department of rural roads	motorcycle	speeding	collision with pedestrian	2	1
10	600643	1/1/2019 0:40	11/1/2019 10:01	Songkhla	department of rural roads	motorcycle	speeding	rear-end collision	2	3
11	599105	1/1/2019 0:45	1/1/2019 10:11	Trat	department of rural roads	motorcycle	speeding	rear-end collision	2	0
12	599598	1/1/2019 0:45	4/1/2019 12:01	Lamphun	department of rural roads	motorcycle	speeding	rollover/fallen on straight road	1	0
13	3792828	1/1/2019 0:45	3/3/2020 10:51	Chumphon	department of rural roads	motorcycle	speeding	rollover/fallen on straight road	1	1
14	599939	1/1/2019 1:00	5/1/2019 10:38	Phuket	department of rural roads	motorcycle	speeding	rollover/fallen on straight road	1	0
15	605004	1/1/2019 1:00	29/3/2019 10:01	Saraburi	department of rural roads	motorcycle	speeding	rollover/fallen on straight road	1	0
16	607600	1/1/2019 1:00	10/6/2019 14:33	Ratchaburi	department of rural roads	motorcycle	failure to yield	rear-end collision	2	1
17	599060	1/1/2019 1:04	1/1/2019 11:52	Phra Nakhon Si Thammarat	department of rural roads	4-wheel passenger vehicle	speeding	pedestrian	1	0
18	599599	1/1/2019 1:15	4/1/2019 14:19	Saraburi	department of rural roads	motorcycle	cutting in	head-on collision	2	0
19	599947	1/1/2019 1:20	5/1/2019 11:08	Nakhon Ratchasima	department of rural roads	motorcycle	speeding	rollover/fallen on straight road	1	1
20	606721	1/1/2019 1:20	8/5/2019 14:54	Prachuap Khiri Khan	department of rural roads	motorcycle	speeding	rollover/fallen on straight road	1	0
21	599512	1/1/2019 1:30	4/1/2019 14:23	Saraburi	department of rural roads	motorcycle	speeding	head-on collision	2	1
22	599536	1/1/2019 1:30	4/1/2019 11:56	Lamphun	department of rural roads	private/passenger car	speeding	rollover/fallen on straight road	1	0
23	599626	1/1/2019 1:30	4/1/2019 12:20	Nakhon Si Thammarat	department of rural roads	4-wheel passenger vehicle	speeding	rollover/fallen on straight road	1	0

# อ่าน/นำเข้าข้อมูล "thai\_road\_accident\_practice.csv")

```
> dt <- read.csv("thai_road_accident_practice.csv")
```

```
> str(dt)
```

```
'data.frame': 81735 obs. of 16 variables:
 $ acc_code      : int  571905 3790870 599075 571924 599523 571982 612782 ...
 $ incident_datetime : chr  "1/1/2019 0:00" "1/1/2019 0:03" "1/1/2019 0:05" ...
 $ report_datetime  : chr  "2/1/2019 6:11" "20/2/2020 13:48" "1/1/2019 10:35" ...
 $ province_en     : chr  "Loburi" "Ubon Ratchathani" "Prachuap Khiri Khan" ...
 $ agency          : chr  "department of rural roads" "department of highways" ...
 $ vehicle_type    : chr  "motorcycle" "private/passenger car" "motorcycle" ...
 $ presumed_cause   : chr  "driving under the influence of alcohol" "speeding" ...
 $ accident_type    : chr  "other" "rollover/fallen on straight road" "head-on collision" ...
 $ number_of_vehicles_involved: int  1 1 2 1 1 1 2 2 2 2 ...
 $ number_of_fatalities : int  0 0 1 0 0 0 0 1 3 0 ...
 $ number_of_injuries : int  2 2 0 1 0 2 2 0 0 1 ...
 $ weather_condition : chr  "clear" "clear" "clear" "clear" ...
 $ latitude         : num  15 15.2 12.4 18.6 15.9 ...
 $ longitude        : num  100.9 104.9 99.9 98.8 100.6 ...
 $ road_description : chr  "straight road" "straight road" "wide curve" "straight road" ...
 $ slope_description : chr  "no slope" "no slope" "slope area" "no slope" .
```

a) ให้นักศึกษาเขียนคำอธิบายตัวแปรโดยสังเกตจากข้อมูลของแต่ละตัวแปร

ตัวแปร	คำอธิบาย
acc_code	รหัสอุบัติเหตุ

ตัวแปร	คำอธิบาย
incident_datetime	เวลาเกิดเหตุ(วัน/เดือน/ปี)
report_datetime	เวลารายงาน(วัน/เดือน/ปี)
province_en	ชื่อจังหวัด
agency	เอเจนซี
vehicle_type	ประเภทของยานพาหนะ
presumed_cause	สาเหตุของการเกิดอุบัติเหตุ
accident_type	ประเภทของอุบัติเหตุ
number_of_vehicles_involved	จำนวนของยานพาหนะที่เกิดเหตุ
number_of_fatalities	จำนวนผู้เสียชีวิต
number_of_injuries	จำนวนผู้บาดเจ็บ
weather_condition	สภาพอากาศ
latitude	วัดระยะทางเหนือ-ใต้ของเส้นศูนย์สูตร
longitude	วัดระยะทางตะวันออก-ตะวันตกของเส้นศูนย์สูตร
road_description	คำอธิบายของถนน
slope_description	คำอธิบายความลาดชัน

- b) ให้นำปีที่เกิดเหตุ (incident\_datetime) ของกลุ่มตัวอย่างแต่ละคน ข้อมูลถูกบันทึกในช่วงปีใด แต่ละปีมีจำนวนคนเกิดเหตุกี่ราย โดยแสดงผลเป็นตาราง

คำตอบ (แสดงคำสั่งและผลลัพธ์)
> dt\$inco <- format(as.Date(dt\$incident_datetime,"%d/%m/%Y"),"%Y")
> dt\$inco <- as.numeric(dt\$inco)
> summary(dt\$inco)
> table(dt\$inco)

- c) ให้นำเดือนที่เกิดเหตุ (incident\_datetime) ของกลุ่มตัวอย่างแต่ละคน แต่ละเดือนมีจำนวนคนเกิดเหตุกี่ราย โดยแสดงผลเป็นตาราง

คำตอบ (แสดงคำสั่งและผลลัพธ์)
------------------------------

```

> dt$incm <- format(as.Date(dt$incident_datetime,"%d/%m/%Y"),"%m")
> dt$incm <- as.numeric(dt$incm)
> summary(dt$incm)
> table(dt$incm)

```

- d) ในแต่ละปีส่วนใหญ่จะเกิดเหตุเดือนใด โดยแสดงผลลัพธ์เป็นตาราง

คำตอบ (แสดงคำสั่งและผลลัพธ์)

```

> dt$incident_date <- as.Date(dt$incident_datetime, "%d/%m/%Y")
> dt$y <- format(dt$incident_date, "%Y")
> dt$m <- format(dt$incident_date, "%m")
> event <- table(dt$year, dt$month)
> max_months <- apply(event, 1, function(x) {
+   names(x)[which.max(x)]
+ })
> re_table <- data.frame(year = names(max_months), month = max_months)
> re_table
  year month
2019 2019   04
2020 2020   12
2021 2021   04
2022 2022   04
> |

```

- e) ให้นำช่วงเวลาที่เกิดเหตุ (incident\_datetime) ของกลุ่มตัวอย่างแต่ละคน และแบ่งกลุ่มช่วงเวลาดังแสดงข้างล่างนี้ เก็บในตัวแปรใหม่ชื่อ incetimegrp ให้นักศึกษาตรวจสอบว่าแต่ละช่วงเวลามีจำนวนคนเกิดเหตุกี่ราย โดยแสดงผลลัพธ์เป็นตาราง

1: 06.01-12.00

2: 12.01-18.00

3: 18.01-00.00

4: 00.01-06.00

คำตอบ (แสดงคำสั่งและผลลัพธ์)

```

> dt$inctime <- as.POSIXct(dt$incident_datetime, format = "%d/%m/%Y %H:%M")
> dt$inctime <- format(as.POSIXct(dt$inctime),format = "%H:%M")
> dt$inctime <- gsub("[:]", ".",dt$inctime)

```



```
> dt$inctime <- as.numeric(dt$inctime)
> summary(dt$inctime)
```

- f) ให้นำระยะห่างของวันนับตั้งแต่วันที่เกิดเหตุ (incident\_datetime) จนถึงวันที่รายงานการเกิดเหตุ (report\_datetime) ของกลุ่มตัวอย่างแต่ละคน โดยแสดงผลลัพธ์เป็นตาราง จำนวนวันที่สั้นที่สุดกี่วัน นานที่สุดกี่วัน โดยเฉลี่ยกี่วัน

คำตอบ (แสดงคำสั่งและผลลัพธ์)

```
> dt$incdate <- as.Date(dt$incident_datetime,"%d/%m/%Y")
> dt$rptdate <- as.Date(dt$report_datetime,"%d/%m/%Y")
> dt$peri <- dt$rptdate- dt$incdate
```

```
> dt$inctime <- as.POSIXct(dt$incident_datetime, format = "%d/%m/%Y %H:%M")
> dt$inctime <- format(as.POSIXct(dt$inctime),format = "%H:%M")
> dt$inctime <- gsub("[:]", ".",dt$inctime)
> dt$inctime <- as.numeric(dt$inctime)
> summary(dt$inctime)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.00   8.00   13.10   12.53   17.55   23.59
> dt$incdate <- as.Date(dt$incident_datetime,"%d/%m/%Y")
> dt$rptdate <- as.Date(dt$report_datetime,"%d/%m/%Y")
> dt$peri <- dt$rptdate- dt$incdate
> dt$days_between <- as.numeric(dt$rptdate - dt$incdate)
> summary(dt$days_between)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.00   8.00   49.00   77.12  103.00   651.00
> dt$incdate <- as.Date(dt$incident_datetime, "%d/%m/%Y %H:%M")
> dt$rptdate <- as.Date(dt$report_datetime, "%d/%m/%Y %H:%M")
> dt$peri <- dt$rptdate - dt$incdate
> result_table <- data.frame(
+   incident_datetime = dt$incident_datetime,
+   report_datetime = dt$report_datetime,
+   days_between = as.numeric(dt$peri)
+ )
> summary(result_table$days_between)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.00   8.00   49.00   77.12  103.00   651.00
> |
```

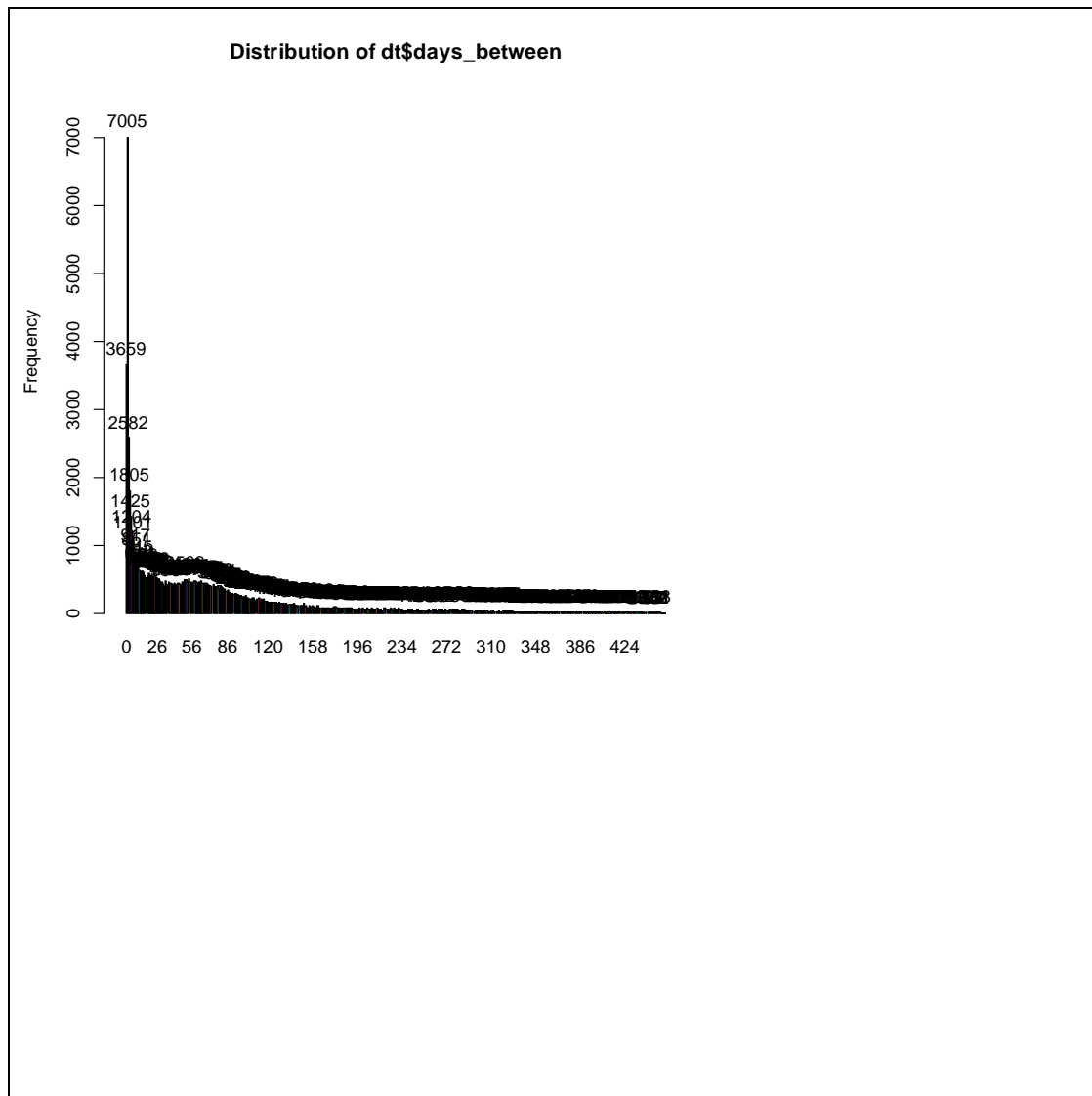
```
dt$inctime <- format(as.Date(dt$incident_datetime,"%d/%m/%Y"),"%H:%M")
dt$inctime <- as.numeric(dt$inctime)
summary(dt$inctime)
table(dt$inctime)
```

```
dt$incm <- format(as.Date(dt$incident_datetime,"%d/%m/%Y"),"%H:%M")
dt$incm <- as.numeric(dt$incm)
summary(dt$incm)
table(dt$incm)
```

```
dt$inctime <- as.POSIXct(dt$incident_datetime, format = "%d/%m/%Y %H:%M")
dt$inctime <- format(as.POSIXct(dt$inctime),format = "%H:%M")
dt$inctime <- gsub("[:]", ".",dt$inctime)
dt$inctime <- as.numeric(dt$inctime)
summary(dt$inctime)
```

```
dt$incdate <- as.Date(dt$incident_datetime,"%d/%m/%Y")
dt$rptdate <- as.Date(dt$report_datetime,"%d/%m/%Y")
```

คำอธิบาย



- g) ให้แสดงข้อมูลที่ผิดปกติในข้อข้างต้น โดยแสดงปี จังหวัด ประเภทของยานพาหนะ จำนวนผู้เสียชีวิต และจำนวนผู้บาดเจ็บ ให้หาระยะห่างของวันนับตั้งแต่วันที่เกิดเหตุ (incident\_datetime) จนถึงวันที่รายงานการเกิดเหตุ

คำตอบ (แสดงคำสั่งและบางส่วน of ผลลัพธ์)

```
> dt$incident_date <- as.Date(dt$incident_datetime, "%d/%m/%Y %H:%M")
> dt$report_date <- as.Date(dt$report_datetime, "%d/%m/%Y %H:%M")
> dt$days_between <- as.numeric(dt$report_date - dt$incident_date)
> un_info <- dt[, c("year", "province_en", "vehicle_type",
+                  "number_of_fatalities", "number_of_injuries", "days_between")]
> un_case <- un_info[un_info$number_of_fatalities >= 5 | un_info$number_of_injuries >= 5]
> head(un_case, 5)
```

year	province_en	vehicle_type	number_of_fatalities
146	2019	Nakhon Ratchasima	4-wheel pickup truck
260	2019	Chiang Rai	private/passenger car
329	2019	Yasothon	4-wheel pickup truck
452	2019	Ubon Ratchathani	passenger pickup truck
681	2019	Pathum Thani	large passenger vehicle

```
number_of_injuries days_between
146 1 10
260 12 1
329 14 9
452 14 1
681 51 424
```

```
dt$incident_date <- as.Date(dt$incident_datetime, "%d/%m/%Y %H:%M")
dt$report_date <- as.Date(dt$report_datetime, "%d/%m/%Y %H:%M")
dt$days_between <- as.numeric(dt$report_date - dt$incident_date)
un_info <- dt[, c("year", "province_en", "vehicle_type",
                  "number_of_fatalities", "number_of_injuries", "days_between")]
un_case <- un_info[un_info$number_of_fatalities >= 5 | un_info$number_of_injuries >= 5]
head(un_case, 5)
```

- h) ให้คำนวณแต่ละปีมีจำนวนผู้บาดเจ็บและเสียชีวิตจากอุบัติเหตุทางถนนโดยเฉลี่ยเท่าไร

คำตอบ (แสดงคำสั่งและผลลัพธ์)

```
> avg_stat <- aggregate(cbind(number_of_fatalities, number_of_injuries)~year, data$
> avg_stat
  year number_of_fatalities number_of_injuries
1 2019      0.1679171      0.9202355
2 2020      0.1564697      0.8273323
3 2021      0.1332551      0.7385247
4 2022      0.1262362      0.7687809
> |
```

- i) ให้คำนวณแต่ละเดือนมีจำนวนผู้บาดเจ็บและเสียชีวิตจากอุบัติเหตุทางถนนโดยเฉลี่ยเท่าไร

คำตอบ (แสดงคำสั่งและผลลัพธ์)

```
> avg_stat <- aggregate(cbind(number_of_fatalities, number_of_injuries)~year+month$
> avg_stat
  year month number_of_fatalities number_of_injuries
1 2019 01      0.1705385      0.8986272
2 2020 01      0.1689291      0.9964806
3 2021 01      0.1671041      0.8578303
4 2022 01      0.1278228      0.8815509
5 2019 02      0.1702290      0.8251908
6 2020 02      0.1872636      0.9413619
7 2021 02      0.1414868      0.7637890
8 2022 02      0.1248295      0.6841746
9 2019 03      0.1462295      0.9022951
10 2020 03      0.1670854      0.8046482
11 2021 03      0.1526882      0.7198925
12 2022 03      0.1159895      0.6376147
13 2019 04      0.1675432      1.0604808
14 2020 04      0.1377119      0.7302260
15 2021 04      0.1135078      0.8422914
16 2022 04      0.1396361      0.8848892
17 2019 05      0.1448382      0.9306626
18 2020 05      0.1596278      0.7136722
19 2021 05      0.1275691      0.6583993
```

- j) ให้หาในแต่ละปีโดยส่วนใหญยานพาหนะประเภทใดที่ประสบอุบัติเหตุทางถนน

คำตอบ (แสดงคำสั่งและผลลัพธ์)

```
> dt <- data.frame(
+   year = c(2019, 2019, 2020, 2020, 2020, 2021, 2021),
+   vehicle_type = c("motorcycle", "car", "car", "motorcycle", "car", "motorcycle")
+ )
> accidents_by_type <- table(dt$year, dt$vehicle_type)
> most_frequent_types <- apply(accidents_by_type, 1, function(x) {
+   names(x)[which.max(x)]
+ })
> re_tab <- data.frame(year=names(most_frequent_types), most_frequent_vehicle_type$
> re_tab
  year most_frequent_vehicle_type
2019 2019                      car
2020 2020                      car
2021 2021                    motorcycle
> |
```

4. ให้นักศึกษาฝึกสำรวจข้อมูลโดยใช้ชุดข้อมูล "CBR64-Nara1\_practice.xlsx" และ "CBR64-Nara2\_practice.xlsx" มีคำอธิบายตัวแปรดังนี้

CBR64-Nara1\_practice.xlsx

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	tit	sex	age	britle	stt	tnb	prv	edu	occ	occ2	mem	mem2	dept	deptfror	depamo	inc31m	inc32m	inc33y	inc53m	inc55y	
2	1) นาย	1) ชาย	49	210125	2) สมรส นานาค	นราธิวาส 1) ประด	ดักปลาไหล				6	6	2) มี	(กรรมารชบว: 50000	12000					10000	
3	3) นาง	2) หญิง	71	104249	4) หม้าย นานาค	นราธิวาส 1) ประด	รับจ้างขายของคค				3	3	1) ไม่มี			7500			1000		
4	3) นาง	2) หญิง	89	2475	4) หม้าย นานาค	นราธิวาส 1) ประด	รับจ้างขายของคค				5	5	2) มี	(กร ช้อรคค	3000	4500			6000		
5	3) นาง	2) หญิง	65	200524	2) สมรส นานาค	นราธิวาส 1) ประด	รับจ้างขายของ				7	7	1) ไม่มี			7500			2000		
6	3) นาง	2) หญิง	50	210225	4) หม้าย นานาค	นราธิวาส 1) ประด	รับจ้างขายของคค				5	5	1) ไม่มี			1000				30000	
7	3) นาง	2) หญิง	75	2489	4) หม้าย นานาค	นราธิวาส 1) ประด	รับจ้างขายของคค				4	4	1) ไม่มี			4500					
8	3) นาง	2) หญิง	48	020525	4) หม้าย นานาค	นราธิวาส 1) ประด	รับจ้างทำขนม				5	5	1) ไม่มี			12000		1000			
9	3) นาง	2) หญิง	42	110825	5) แยก นานาค	นราธิวาส 1) ประด	รับจ้างกริตยง				5	5	1) ไม่มี			7200		750	3000		
10	1) นาย	1) ชาย	55	290625	2) สมรส นานาค	นราธิวาส 1) ประด	เลี้ยงวัว	เลี้ยงสัตว์			8	8	1) ไม่มี			15000		5000		25000	
11	1) นาย	1) ชาย	44	1E+07	2) สมรส นานาค	นราธิวาส 1) ประด	ว่างงาน				5	5	2) มี	(กร สร้างบนำ	30000	3000					
12	3) นาง	2) หญิง	44	2E+07	2) สมรส ไซซิด	นราธิวาส 1) ประด	ว่างงาน				4	4	1) ไม่มี								
13	นางส 2) หญิง		40	210125	2) สมรส นานาค	นราธิวาส 1) ประด	รับจ้างปรุงอาหาร				6	6	2) มี	(กรรมารชบว: 30000	4000				5000		
14	3) นาง	2) หญิง	57	110725	2) สมรส นานาค	นราธิวาส 1) ประด	ว่างงาน				5	5	2) มี	(กรรมารชบว: 30000	14000				1000		
15	นางส 2) หญิง		45	110825	5) แยก นานาค	นราธิวาส 2) มัธยม	คคจากขายข				5	5	2) มี	(กรรมารชบว: 30000	9500				4000		
16	นางส 2) หญิง		37	903252	2) สมรส นานาค	นราธิวาส 1) ประด	รับจ้างขายคคคค				5	5	2) มี	(กรรมารชบว: 30000	7500				1000		
17	นางส 2) หญิง		50	260625	2) สมรส นานาค	นราธิวาส 1) ประด	เลี้ยงแกะ				4	4	2) มี	(กร สร้างแล	30000	7000			2000		
18	นางส 2) หญิง		33	010525	2) สมรส นานาค	นราธิวาส 2) มัธยม	รับจ้างกริตยง				4	4	2) มี	(กรรมารชบว: 30000	7000				3000		
19	1) นาย	1) ชาย	54	220525	2) สมรส ไซซิด	นราธิวาส 1) ประด	รับจ้างกริตยง				5	5	1) ไม่มี			6000			3000		
20	นาง	2) หญิง	52	180525	2) สมรส ไซซิด	นราธิวาส 1) ประด	รับจ้างกริตยง				5	5	1) ไม่มี			75000			2000		
21	นางส 2) หญิง		59	230625	4) หม้าย ไซซิด	นราธิวาส 1) ประด	ว่างงาน				3	3	1) ไม่มี			10000					
22	1) นาย	1) ชาย	58	1E+07	2) สมรส นานาค	นราธิวาส 1) ประด	กริตยง	เลี้ยง เป็			3	3	1) ไม่มี			7000				10000	
23	นางส 2) หญิง		58	2E+07	2) สมรส นานาค	นราธิวาส 1) ประด	ว่างงาน				7	7	1) ไม่มี			5000				25000	
24	1) นาย	1) ชาย	49	1E+07	2) สมรส นานาค	นราธิวาส 1) ประด	รับจ้างคค	เลี้ยงวัว			3	3	1) ไม่มี			5000			3000		
25	1) นาย	1) ชาย	53	1E+07	2) สมรส นานาค	ตากใบ	1) ประด	รับจ้างคค	เลี้ยงวัว		6	6	1) ไม่มี			1300				30000	
26	1) นาย	1) ชาย	37	081225	2) สมรส นานาค	นราธิวาส 2) มัธยม	รับจ้างคค	เลี้ยงแพ			6	6	1) ไม่มี			7000		1000			

CBR64-Nara2\_practice.xlsx

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	tit	sex	stt	brtde	tmb	prv	edu	occ	mem	covid	expend	inc53m	inc55y	inc57r	inc58r
2	นางส	หญิง	สมรส	260425C	มาโง	นราธิวาส	1) ประภ	กริดยาง		6	2) มีงาน	1) ค่าอา	3000		
3	นางส	หญิง	หม้าย	1012498	มาโง	นราธิวาส	ไม่เรียน	ทำสวน		5	2) มีงาน	1) ค่าไฟ	300 บาท	5000	
4	นางส	หญิง	หม้าย	171025C	มาโง	นราธิวาส	ไม่เรียน	ทำสวน		4	2) มีงาน	1) ค่าไฟ	200บาท/เดือน	2)	24 300
5	นางส	หญิง	หม้าย	2112528	มาโง	นราธิวาส	2) มัธยม	ทำสวน		4	2) มีงาน	1) ค่าไฟ	650 บาท/เดือน	2)	ค่าอาหาร 100 บาท
6	นาง	หญิง	สมรส	2102528	มาโง	นราธิวาส	2) มัธยม	ทำสวน		7	2) มีงาน	1) ค่าไฟ	4000		
7	นางส	หญิง	สมรส	604253	มาโง	นราธิวาส	2) มัธยม	กริดยาง		5	2) มีงาน	1) ค่าอาหาร	3 มื้อ	200 บาท	10 1900
8	นาง	หญิง	สมรส	3105253	มาโง	นราธิวาส	1) ประภ	กริดยาง		4	2) มีงาน	1) ค่าอาหาร	3 มื้อ	200 บาท	22 2000
9	นาย	ชาย	สมรส	28/07/25	มาโง	นราธิวาส	1) ประภ	ทำการเค		5	2) มีงาน	1) ค่าอา	9000		
10	นางส	หญิง	สมรส	1906251	มาโง	นราธิวาส	1) ประภ	เกษตร		5	2) มีงาน	1) ค่าอา	4000		
11	นาย	ชาย	โสด	7062507	มาโง	นราธิวาส	1) ประภ	กริดยาง		2	2) มีงาน	1) ค่าอา	2000		
12	นางส	หญิง	สมรส	1605252	มาโง	นราธิวาส	2) มัธยม	ทำสวน		4	2) มีงาน	1) ค่าอา	3000		
13	นางส	หญิง	สมรส	2507251	มาโง	นราธิวาส	ไม่เรียน	ลูกจ้างที่		16	2) มีงาน	1) ค่าไฟ/ค่าไฟ	650 บาท/เดือน	2)	ค่าอาหาร 100
14	นางส	หญิง	สมรส	1309254	มาโง	นราธิวาส	2) มัธยม	กริดยาง		3	2) มีงาน	1) ค่าอาหาร	3 มื้อ	200/วัน	2 15 2900
15	นางส	หญิง	สมรส	3103253	มาโง	นราธิวาส	2) มัธยม	แม่บ้าน		5	2) มีงาน	1) ค่าอาหาร	3 มื้อ	300 บาท/วัน	2) ค่าไฟ 20 บาท
16	นาย	ชาย	สมรส	0604252	มาโง	นราธิวาส	2) มัธยม	กริดยาง		5	1) ไม่ได้	1) ค่าอา	3000		
17	นางส	หญิง	สมรส	308253C	มาโง	นราธิวาส	1) ประภ	เกษตร		4	2) มีงาน	1) ค่าอาหาร	3 มื้อ	300 บาท	3 3000
18	นาย	ชาย	สมรส	1106252	มาโง	นราธิวาส	ไม่ไดเรีย	กริดยาง		4	1) ไม่ได้	1) ค่าไฟ	3000		
19	นาย	ชาย	สมรส	1005253	มาโง	นราธิวาส	2) มัธยม	รับจ้างทั่ว		9	2) มีงาน	1) ค่าชดเชย	นักเรียน	900 บาท/ปี	
20	นาย	ชาย	สมรส	1006251	มาโง	นราธิวาส	2) มัธยม	ทำสวน		6	1) ไม่ได้	1) ค่าไฟ	400 บาท/เดือน	2)	6000
21	นาย	ชาย	สมรส	1012494	มาโง	นราธิวาส	กรณ	ทำสวน		4	2) มีงาน	1) ค่าอา	4000		
22	นาย	ชาย	สมรส	3052527	มาโง	นราธิวาส	1) ประภ	ลูกจ้างที่		5	2) มีงาน	1) ค่าอาหาร	100 บาท/วัน	2)	ค่าไฟ 300 บาท
23	นางส	หญิง	สมรส	3062541	มาโง	นราธิวาส	2) มัธยม	ลูกจ้างทั่ว		3	2) มีงาน	1) ค่าอาหาร	3 มื้อ	300/วัน	2) ค่าไฟ 20 บาท/เดือน

- a) ให้ตรวจสอบข้อมูล CBR64-Nara1\_practice.xlsx และ CBR64-Nara2\_practice.xlsx มีจำนวนกี่แถวที่คอดัมน์

คำตอบ (แสดงคำสั่งและผลลัพธ์)

```

R Console
> nr_1 <- read_excel("CBR64-Nara1_practice.xlsx")
> nr_1 <- as.data.frame(nr_1)
> str(nr_1)
'data.frame':   40 obs. of  20 variables:
 $ tit      : chr  "1) นาย" "3) นาง" "3) นาง" ...
 $ sex      : chr  "1) ชาย" "2) หญิง" "2) หญิง" ...
 $ age      : num  49 71 89 65 50 75 48 42 55 44 ...
 $ brtde    : chr  "21012515" "1042493" "2475" "20052499" ...
 $ stt      : chr  "2) สมรส" "4) หม้าย" "4) หม้าย" "2) สมรส" ...
 $ tmb      : chr  "นายพล" "นายพล" "นายพล" "นายพล" ...
 $ prv      : chr  "นราธิวาส" "นราธิวาส" "นราธิวาส" "นราธิวาส" ...
 $ edu      : chr  "1) ประถมศึกษา" "ไม่ไดเรียน" "ไม่ไดเรียน" "1) ประถมศึกษา" ...
 $ occ      : chr  "ลูกจ้างทั่วไป" "รับจ้างของหลวง" "รับจ้าง" "รับจ้าง" ...
 $ occ2     : chr  NA NA NA NA ...
 $ mem      : num  6 3 5 7 5 4 5 5 8 5 ...
 $ mem2     : num  6 3 5 7 5 4 5 5 8 5 ...
 $ dept     : chr  "2) มี (ทุกหน่วยวัดเป็นจำนวนไร่ เช่น ไร่ในไร่ในไร่)" ...
 $ deptfrom : chr  NA NA "ใช้หน่วยวัดไร่" NA ...
 $ depamount: num  50000 NA 3000 NA NA NA NA NA 30000 ...
 $ inc31m   : num  12000 7500 4500 7500 1000 4500 12000 7200 15000 3000 ...
 $ inc32m   : logi  NA NA NA NA NA NA ...
 $ inc33y   : num  NA NA NA NA NA NA 1000 750 5000 NA ...
 $ inc53m   : num  NA 1000 6000 2000 NA NA NA 3000 NA NA ...
 $ inc55y   : num  10000 NA NA NA 30000 NA NA NA 25000 NA ...
> nr_2 <- read_excel("CBR64-Nara2_practice.xlsx")
> nr_2 <- as.data.frame(nr_2)
> str(nr_2)
'data.frame':   40 obs. of  15 variables:

```

- b) ให้ทำการรวมข้อมูล CBR64-Nara1\_practice.xlsx และ CBR64-Nara2\_practice.xlsx หลังจากรวมแล้วข้อมูลมีจำนวนกี่แถวกี่คอลัมน์ (รวมทางแถว)

คำตอบ (แสดงคำสั่งและผลลัพธ์)

```

> library(tidyverse)
> inc55y <- read_csv("inc55y.csv")
> nr_3 <- rbind(nr_11, nr_22)
> des(nr_3)

No. of observations = 80
Variable      Class      Description
1 tit          character
2 sex          character
3 stt          character
4 brtdte       character
5 tmb          character
6 prv          character
7 edu          character
8 occ          character
9 mem          numeric
10 inc53m      numeric
11 inc55y      numeric

```

- c) กลุ่มตัวอย่างมีอายุเฉลี่ยเท่าไร มากที่สุดเท่าไร และน้อยที่สุดเท่าไร

คำตอบ (แสดงคำสั่งและผลลัพธ์)

```

44 2) นามสกุล 2) หมู่ 4) หมู่ 2112538 อำเภอ นารายณ์
45 3) นามสกุล 2) หมู่ 2) หมู่ 2102538 อำเภอ นารายณ์
60 1) นามสกุล 2) หมู่ 1012494 อำเภอ นารายณ์
74 1) นามสกุล 2) หมู่ 1102533 อำเภอ นารายณ์
76 3) นามสกุล 2) หมู่ 1012495 อำเภอ นารายณ์

```

```

occ mem inc53m inc55y
39 1) ปริญญาโท <NA> 4 NA 20000
40 ไม่เรียนหนังสือ จบจากโรงเรียน 7 5000 NA
42 ไม่เรียนหนังสือ จบจากโรงเรียน 5 NA 5000
44 2) มัธยมศึกษาตอนต้น/ตอนปลาย จบจาก 4 NA NA
45 2) มัธยมศึกษาตอนต้น/ตอนปลาย จบจาก 7 4000 NA
60 คน จบจาก โรงเรียน 4 4000 NA
74 2) มัธยมศึกษาตอนต้น/ตอนปลาย จบจาก 5 NA NA
76 ไม่เรียนหนังสือ จบจาก 3 2000 NA

```

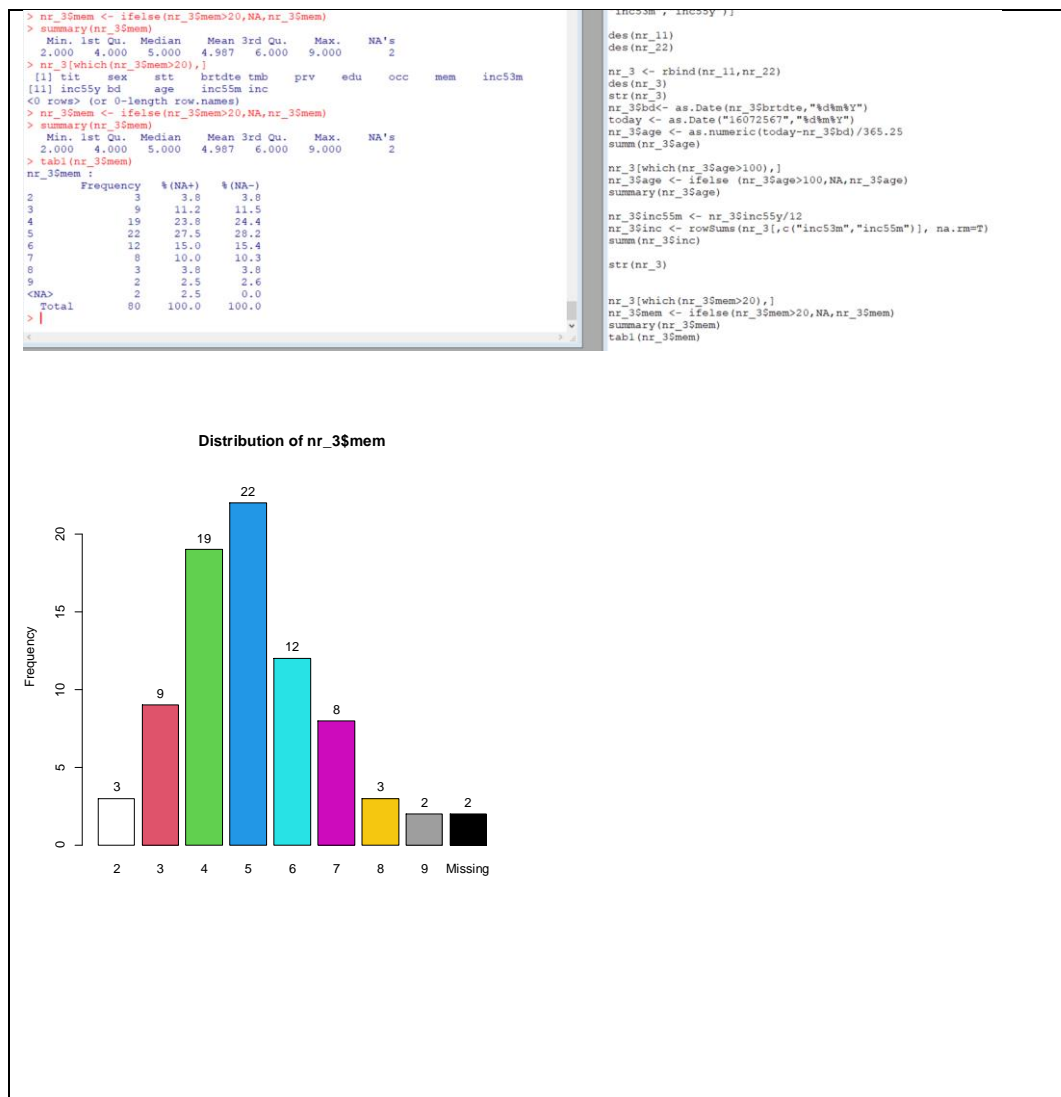
```

bd age
39 0523-12-10 2043.556
40 0527-12-10 2039.556
42 0498-12-10 2068.553
44 0528-12-21 2038.524
45 0528-02-21 2039.357
60 0494-12-10 2072.553
74 0533-02-11 2034.382
76 0495-12-10 2071.554
> nr_3$age <- ifelse (nr_3$age>100, NA, nr_3$age)
> summary(nr_3$age)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
 24.34  41.24   50.15   48.72   56.59   68.15     21

```

- d) กลุ่มตัวอย่างมีจำนวนสมาชิกในบ้านโดยเฉลี่ยเท่าไร มากที่สุดเท่าไร และน้อยที่สุดเท่าไร

คำตอบ (แสดงคำสั่งและผลลัพธ์)



e) กลุ่มตัวอย่างมีรายได้ต่อเดือนโดยเฉลี่ยเท่าไร มากที่สุดเท่าไร และน้อยที่สุดเท่าไร

คำตอบ (แสดงคำสั่งและผลลัพธ์)

```
> nr_3$inc <- rowSums(nr_3[,c("inc53m")], na.rm=T)
Error: unexpected symbol in "nr_3$inc <- rowSums(nr_3[,c("inc53m"
> sum(nr_3$inc)
Error in order(x) : argument 1 is not a vector
> nr_3$inc55m <- nr_3$inc55y/12
> nr_3$inc <- rowSums(nr_3[,c("inc53m")], na.rm=T)
Error: unexpected symbol in "nr_3$inc <- rowSums(nr_3[,c("inc53m"
> sum(nr_3$inc)
Error in order(x) : argument 1 is not a vector
> nr_3$inc55m <- nr_3$inc55y/12
> nr_3$inc <- rowSums(nr_3[,c("inc53m")], na.rm=T)
Error: unexpected symbol in "nr_3$inc <- rowSums(nr_3[,c("inc53m"
> nr_3$inc55m <- nr_3$inc55y/12
> nr_3$inc <- rowSums(nr_3[,c("inc53m")], na.rm=T)
Error in rowSums(nr_3[, c("inc53m")], na.rm = T) :
'x' must be an array of at least two dimensions
> sum(nr_3$inc)
Error in order(x) : argument 1 is not a vector
> nr_3$inc55m <- nr_3$inc55y/12
> nr_3$inc <- rowSums(nr_3[,c("inc53m")], na.rm=T)
Error in rowSums(nr_3[, c("inc53m")], na.rm = T) :
'x' must be an array of at least two dimensions
> nr_3$inc55m <- nr_3$inc55y/12
> nr_3$inc <- rowSums(nr_3[,c("inc53m","inc55m")], na.rm=T)
> sum(nr_3$inc)
  obs. mean   median    s.d.    min.    max.
80 2219.375 2000    2880.147 0    20000
> |
```

```
nr_22 <- nr_2[,c("tit","sex","stt",
"brtde","tmb","prv","edu","occ","mem",
"inc53m","inc55y")]
des(nr_11)
des(nr_22)
nr_3 <- rbind(nr_11,nr_22)
des(nr_3)
str(nr_3)
nr_3$bdc <- as.Date(nr_3$brtde,"%d%m%y")
today <- as.Date("16072567","%d%m%y")
nr_3$age <- as.numeric(today-nr_3$bdc)/365.25
sum(nr_3$age)
nr_3[which(nr_3$age>100),]
nr_3$age <- ifelse (nr_3$age>100,NA,nr_3$age)
summary(nr_3$age)
nr_3$inc55m <- nr_3$inc55y/12
nr_3$inc <- rowSums(nr_3[,c("inc53m","inc55m")], na.rm=T)
sum(nr_3$inc)
str(nr_3)
nr_3[which(nr_3$mem>20),]
nr_3$mem <- ifelse(nr_3$mem>20,NA,nr_3$mem)
summary(nr_3$mem)
tabi(nr_3$mem)
```



