506 HW1

Fang Zhou

September 30, 2018

Q2

```
source('ps1_q2.R')
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
## Warning: Column `carrier` joining character vector and factor, coercing
## into character vector
```

```
source('ps1_q3.R')
```

a

The airlines that were responsible for at least 1% of the flights departing any of the three NYC airports between January 1 and October 31, 2013:

```
cap_title1 = '**Table 1.** *Airlines responsible for at least 1% between January 1 and October 31.*'
knitr::kable(al13_at_least_0.01,digit=4,caption=cap_title1)
```

Table 1. *Airlines responsible for at least 1% between January 1 and October 31.*

carrier	percent
9E	0.0541
AA	0.0975
B6	0.1621
DL	0.1428
EV	0.1613
FL	0.0101
MQ	0.0789
UA	0.1737

US	0.0612
VX	0.0151
WN	0.0360

b

```
cap_title2 = '**Table 2.** *Percent and change of annual flights in 2013&2014.
*'
knitr::kable(AL_13_14,digit=3,caption = cap_title2)
```

Table 2. *Percent and change of annual flights in 2013&2014.*

carrier	percent_13	percent_14	change	CI_13_Left	CI_13_Right	CI_14_Left	CI_14_Right	change_Left	change_Right
Endeavor Air Inc.	0.054	0.000	-0.054	0.053	0.055	0.000	0.000	-0.055	-0.053
American Airlines Inc.	0.098	0.104	0.006	0.096	0.099	0.103	0.105	0.005	0.008
JetBlue Airways	0.162	0.176	0.014	0.161	0.163	0.174	0.177	0.011	0.016
Delta Air Lines Inc.	0.143	0.165	0.022	0.141	0.144	0.163	0.166	0.020	0.024
ExpressJet Airlines Inc.	0.161	0.157	-0.004	0.160	0.163	0.156	0.159	-0.006	-0.002
AirTran Airways Corporation	0.010	0.005	-0.005	0.010	0.010	0.005	0.005	-0.006	-0.005
Envoy Air	0.079	0.073	-0.006	0.078	0.080	0.072	0.074	-0.007	-0.004
United Air Lines Inc.	0.174	0.183	0.009	0.172	0.175	0.181	0.184	0.007	0.011
US Airways Inc.	0.061	0.066	0.005	0.060	0.062	0.065	0.067	0.004	0.006
Virgin America	0.015	0.019	0.004	0.015	0.016	0.018	0.019	0.003	0.005
Southwest Airlines Co.	0.036	0.047	0.011	0.035	0.037	0.046	0.048	0.010	0.012

Airline shows the largest increase:

```
al_13_14[which.max(al_13_14$change),1]
```

```
## # A tibble: 1 x 1
##   carrier
##   <chr>
## 1 Delta Air Lines Inc.
```

Airline shows the largest decrease:

```
al_13_14[which.min(al_13_14$change),1]
```

```
## # A tibble: 1 x 1
##   carrier
##   <chr>
## 1 Endeavor Air Inc.
```

Why do some airlines show an increase in the percent of flights but a decrease in the number of flights?

Because in 2014, the Endeavor Air Inc. didn't have any flights at NYC airport. There are only 10 carriers left and thus some airlines have larger proportion than 2013 although they have fewer flights.

c

```
cap_title3='**Table 3.** *Percents and CIs of flights each airline is respons
ible for at three NYC airports.*'
knitr::kable(FL_13_14,digit=4,caption = cap_title3)
```

Table 3. *Percents and CIs of flights each airline is responsible for at three NYC airports.*

carrier	airport	percent	CI_left	CI_right
9E	EWR	0.0061	0.0058	0.0064
AA	EWR	0.0295	0.0287	0.0302
B6	EWR	0.0578	0.0568	0.0588
DL	EWR	0.0408	0.0399	0.0416
EV	EWR	0.3470	0.3450	0.3491
MQ	EWR	0.0117	0.0113	0.0122
UA	EWR	0.3948	0.3927	0.3969
US	EWR	0.0379	0.0370	0.0387
VX	EWR	0.0154	0.0149	0.0159
WN	EWR	0.0528	0.0518	0.0538
9E	JFK	0.0760	0.0748	0.0772
AA	JFK	0.1334	0.1318	0.1349
B6	JFK	0.3958	0.3936	0.3980

DL	JFK	0.2052	0.2034	0.2070
EV	JFK	0.0129	0.0123	0.0134
MQ	JFK	0.0656	0.0645	0.0667
UA	JFK	0.0439	0.0430	0.0448
US	JFK	0.0293	0.0285	0.0300
VX	JFK	0.0349	0.0341	0.0358
9E	LGA	0.0134	0.0129	0.0140
AA	LGA	0.1438	0.1422	0.1454
B6	LGA	0.0571	0.0560	0.0581
DL	LGA	0.2207	0.2189	0.2226
EV	LGA	0.1018	0.1004	0.1032
FL	LGA	0.0239	0.0232	0.0245
MQ	LGA	0.1580	0.1564	0.1596
UA	LGA	0.0754	0.0742	0.0766
US	LGA	0.1257	0.1242	0.1272
VX	LGA	0.0001	0.0000	0.0001
WN	LGA	0.0697	0.0686	0.0709

Largest carrier at airport EWR:

```
fl_EWR[which.max(fl_EWR$percent.y),1]
```

```
## # A tibble: 1 x 1
##   carrier
##   <chr>
## 1 UA
```

Largest carrier at airport JFK:

```
fl_JFK[which.max(fl_JFK$percent.y),1]
```

```
## # A tibble: 1 x 1
##   carrier
##   <chr>
## 1 B6
```

Largest carrier at airport LGA:

```
fl_LGA[which.max(fl_LGA$percent.y),1]
```

```
## # A tibble: 1 x 1
##   carrier
##   <chr>
## 1 DL
```

Q3

a

Percent of homes have stucco construction as the major outside wall material within each division

```
cap_title4='**Table 4.** *Percent of homes have stucco construction as the ma
jor outside wall material within each division.*'
knitr::kable(div_stucco,digit=4,caption = cap_title4)
```

Table 4. *Percent of homes have stucco construction as the major outside wall material within each division.*

division	percent
1	0.0123
2	0.0206
3	0.0066
4	0.0487
5	0.1062
6	0.0042
7	0.0299
8	0.1661
9	0.6425
10	0.4459

division with highest porportion

```
div_stucco[which.max(div_stucco$percent),1]
```

```
## # A tibble: 1 x 1
##   division
##      <int>
## 1        9
```

division with lowest porportion

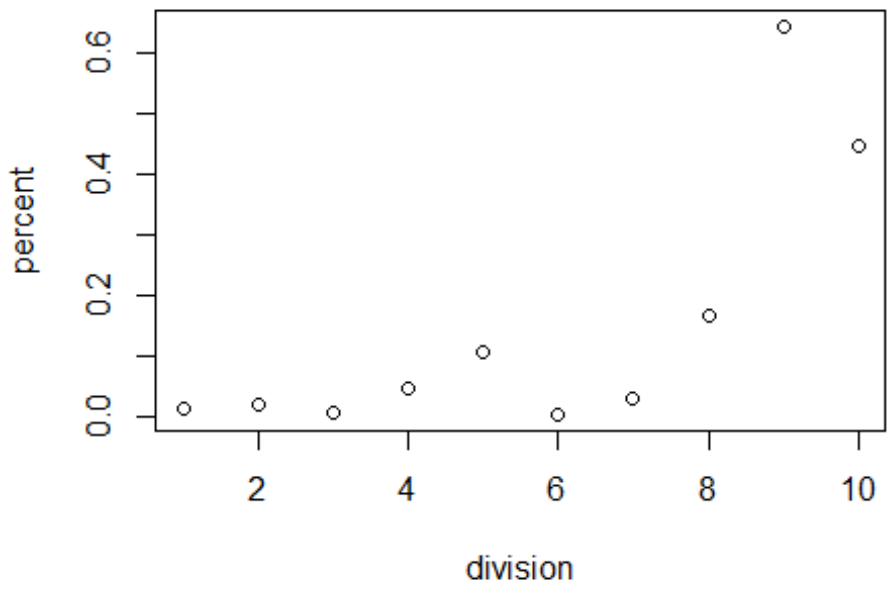
```
div_stucco[which.min(div_stucco$percent),1]
```

```
## # A tibble: 1 x 1
##   division
##      <int>
## 1        6
```

From the plot below we can easily see which divisions have the highest and lowest percent.

```
plot(div_stucco,main='Fig.1 Plot of the percent within each division')
```

Fig.1 Plot of the percent within each division



b

Average total electricity usage in kilowatt hours in each division

```
cap_title5='**Table 5.** *Average total electricity usage in KWH in each divi
sion.*'
knitr::kable(elec_sum,caption = cap_title5)
```

Table 5. *Average total electricity usage in KWH in each division.*

DIVISION	total	average
1	42298294272	7514.561
2	130178972232	8465.442
3	165177599349	9128.663
4	87109342326	10523.828
5	315657431589	13446.621
6	104618118728	14535.969
7	197244103227	14324.259
8	35607816739	8384.471
9	44554730384	10442.017
10	144788716480	8100.405

Average total electricity usage in KWH in each division at urban and rural areas

```
cap_title6='**Table 6.** *Average total electricity usage in KWH in each division stratified by urban and rural status.*'
knitr::kable(elec_UR_sum,caption = cap_title6)
```

Table 6. *Average total electricity usage in KWH in each division stratified by urban and rural status.*

DIVISION	UATYP10	average
1	R	9001.054
1	U	7626.553
2	R	12223.389
2	U	7788.251
3	R	13500.024
3	U	7775.891
4	R	14173.932
4	U	9320.237
5	R	15941.991
5	U	12825.487
6	R	16332.679
6	U	14168.218
7	R	16317.160
7	U	14061.286
8	R	9356.001
8	U	8143.720
9	R	8610.426
9	U	10670.228
10	R	14114.760
10	U	7049.627

c

Division has the largest disparity between urban and rural areas in terms of the proportion of homes with internet access

```
cap_title7='**Table 7.** *Proportion of homes with Internet Access in each division stratified by urban and rural status.*'
knitr::kable(UR_IT,digit=4,caption = cap_title7)
```

Table 7. *Proportion of homes with Internet Access in each division stratified by urban and rural status.*

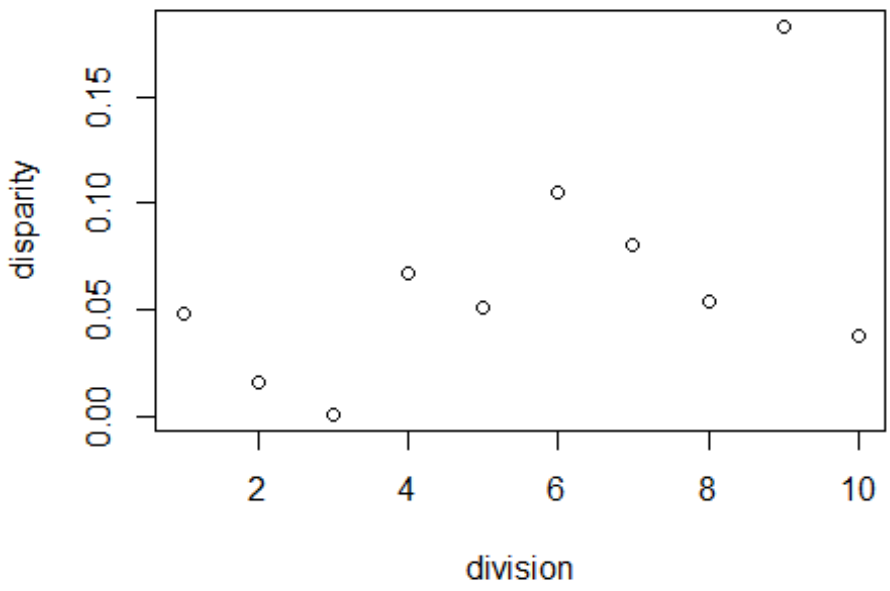
division	urbantype	proportion

1	R	0.8579
1	U	0.9060
2	R	0.9129
2	U	0.8970
3	R	0.8621
3	U	0.8615
4	R	0.8033
4	U	0.8704
5	R	0.8204
5	U	0.8714
6	R	0.6903
6	U	0.7949
7	R	0.7650
7	U	0.8456
8	R	0.8193
8	U	0.8731
9	R	0.6675
9	U	0.8506
10	R	0.8528
10	U	0.8905

Plot of the disparity between urban and rural areas in different divisions

```
plot(abs(disp),main='Fig.2 Disparity between urban and rural areas',xlab='div
ision',ylab='disparity')
```

Fig.2 Disparity between urban and rural areas



From Fig.2 which shows the disparity in different divisions we can easily see that Divison 9 has the largest disparity between urban and rural areas. We can use R code to test it:

```
which.max(abs(disp))
```

```
## [1] 9
```