

Problem Set 1

Runwei Wang

Question 2

Question 2 considers data of flights originating in New York City, NY (NYC) in 2013 and 2014. Data of 2013 from package `nycflights13`, whose author is Hadley Wickham.

Here uses `source()` to execute R files.

```
source("ps1_q2.R")
```

a.

The chart below shows the airlines that takes up at least 1% of the flights departing any of the three NYC airports between January 1 and October 31.

```
knitr::kable(airlines_more_than_onepercent)
```

carrier	percent
Endeavor Air Inc.	0.0541345
American Airlines Inc.	0.0975467
JetBlue Airways	0.1620802
Delta Air Lines Inc.	0.1427571
ExpressJet Airlines Inc.	0.1613339
AirTran Airways Corporation	0.0101111
Envoy Air	0.0789059
United Air Lines Inc.	0.1737196
US Airways Inc.	0.0612425
Virgin America	0.0150512
Southwest Airlines Co.	0.0360482

b.

The chart below compares the number and percent of annual flights in the first 10 months of 2013 and the first 10 months of 2014.

```
knitr::kable(change)
```

Carrier	Percent_13	CI_13	Percent_14	CI_14	Change
Endeavor Air Inc.	0.0541345	(NA, NA)	NA	(0.0533, 0.0550)	NA
American Airlines Inc.	0.0975467	(0.1026, 0.1050)	0.1038308	(0.0965, 0.0986)	0.0062841
JetBlue Airways	0.1620802	(0.1741, 0.1771)	0.1755870	(0.1607, 0.1634)	0.0135068
Delta Air Lines Inc.	0.1427571	(0.1631, 0.1660)	0.1645494	(0.1415, 0.1440)	0.0217923
ExpressJet Airlines Inc.	0.1613339	(0.1558, 0.1586)	0.1571910	(0.1600, 0.1627)	-0.0041429
AirTran Airways Corporation	0.0101111	(0.0047, 0.0052)	0.0049385	(0.0097, 0.0105)	-0.0051726
Envoy Air	0.0789059	(0.0722, 0.0743)	0.0732642	(0.0779, 0.0799)	-0.0056417
United Air Lines Inc.	0.1737196	(0.1811, 0.1842)	0.1826454	(0.1723, 0.1751)	0.0089258
US Airways Inc.	0.0612425	(0.0652, 0.0671)	0.0661229	(0.0604, 0.0621)	0.0048804
Virgin America	0.0150512	(0.0184, 0.0195)	0.0189368	(0.0146, 0.0155)	0.0038856
Southwest Airlines Co.	0.0360482	(0.0462, 0.0478)	0.0469848	(0.0354, 0.0367)	0.0109366

Among which, Delta Air Lines Inc. has the largest positive change, which is 2.18%.

```
knitr::kable(which.max(change$Change))
```

```
—
x
—
4
—
```

Besides, Envoy Air has the largest negative change, which is -0.565%.

```
knitr::kable(which.min(change$Change))
```

```
—
x
—
7
—
```

c.

United Air Lines Inc. is the largest carrier in EWR.

```
knitr::kable(which.max(ewr))
```

```
—
x
—
20
—
```

JetBlue Airways is the largest carrier in JFK.

```
knitr::kable(which.max(jfk))
```

```
—
x
—
8
—
```

Delta Air Lines Inc. is the largest carrier in LGA.

```
knitr::kable(which.max(lga))
```

```
—
x
—
12
—
```

The chart shows the percent of flights each airline is responsible for among NYC airports.

```
knitr::kable(flight_airport_11)
```

Carrier	Origin	Count	Percent	CI
Endeavor Air Inc.	EWR	1268	0.0060893	(0.0059, 0.0063)
Endeavor Air Inc.	JFK	14651	0.0760056	(0.0753, 0.0767)
Endeavor Air Inc.	LGA	2541	0.0134377	(0.0131, 0.0137)
American Airlines Inc.	EWR	6136	0.0294667	(0.0290, 0.0299)
American Airlines Inc.	JFK	25706	0.1333562	(0.1325, 0.1342)
American Airlines Inc.	LGA	27189	0.1437849	(0.1429, 0.1447)
JetBlue Airways	EWR	12030	0.0577713	(0.0572, 0.0584)

Carrier	Origin	Count	Percent	CI
JetBlue Airways	JFK	76296	0.3958042	(0.3946, 0.3971)
JetBlue Airways	LGA	10788	0.0570507	(0.0565, 0.0576)
Delta Air Lines Inc.	EWR	8495	0.0407953	(0.0403, 0.0413)
Delta Air Lines Inc.	JFK	39561	0.2052324	(0.2042, 0.2063)
Delta Air Lines Inc.	LGA	41737	0.2207197	(0.2197, 0.2218)
ExpressJet Airlines Inc.	EWR	72267	0.3470454	(0.3458, 0.3483)
ExpressJet Airlines Inc.	JFK	2477	0.0128500	(0.0126, 0.0131)
ExpressJet Airlines Inc.	LGA	19248	0.1017901	(0.1010, 0.1026)
AirTran Airways Corporation	LGA	4511	0.0238557	(0.0235, 0.0242)
Envoy Air	EWR	2443	0.0117319	(0.0115, 0.0120)
Envoy Air	JFK	12637	0.0655575	(0.0649, 0.0662)
Envoy Air	LGA	29876	0.1579947	(0.1571, 0.1589)
United Air Lines Inc.	EWR	82211	0.3947991	(0.3936, 0.3960)
United Air Lines Inc.	JFK	8458	0.0438779	(0.0434, 0.0444)
United Air Lines Inc.	LGA	14263	0.0754277	(0.0748, 0.0761)
US Airways Inc.	EWR	7885	0.0378659	(0.0374, 0.0384)
US Airways Inc.	JFK	5640	0.0292589	(0.0288, 0.0297)
US Airways Inc.	LGA	23761	0.1256564	(0.1248, 0.1265)
Virgin America	EWR	3210	0.0154153	(0.0151, 0.0157)
Virgin America	JFK	6734	0.0349343	(0.0345, 0.0354)
Virgin America	LGA	15	0.0000793	(0.0001, 0.0001)
Southwest Airlines Co.	EWR	10995	0.0528009	(0.0522, 0.0534)
Southwest Airlines Co.	LGA	13182	0.0697110	(0.0691, 0.0704)

Question 3

Question 3 concerns about data from 2015 Residential Energy Consumption Survey, which can be found from U.S. Energy Information Administration.

Here uses `source()` to execute R files.

```
source("ps1_q3.R")
```

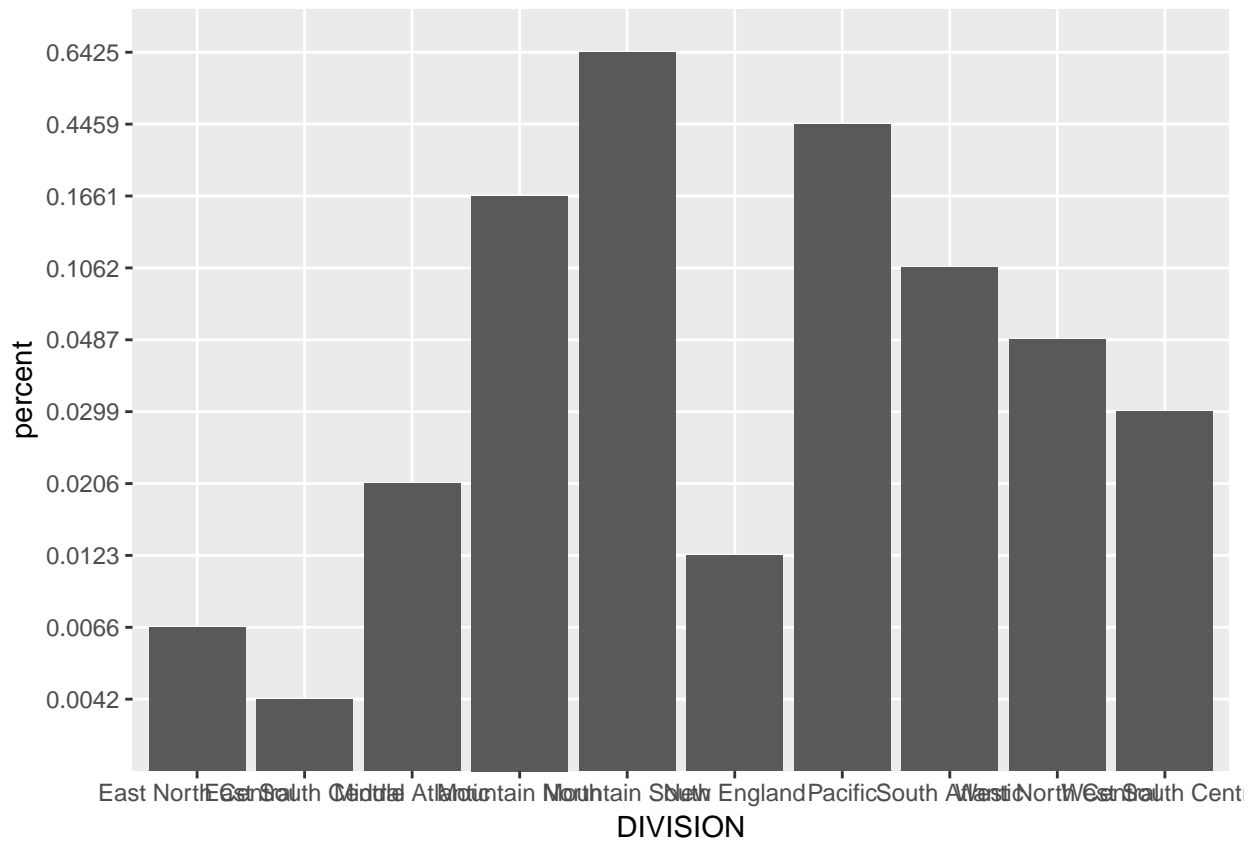
a.

The chart shows the percent of homes have stucco construction as the major outside wall material within each division.

```
knitr::kable(stucco_division[, -1])
```

DIVISION	sum	percent
New England	69040.30	0.0123
Middle Atlantic	317021.99	0.0206
East North Central	118796.73	0.0066
West North Central	402740.72	0.0487
South Atlantic	2492959.08	0.1062
East South Central	30444.77	0.0042
West South Central	411352.10	0.0299
Mountain North	705216.82	0.1661
Mountain South	2741373.16	0.6425
Pacific	7969431.30	0.4459

```
ggplot(stucco_division,aes(x=DIVISION,y=percent))+geom_bar(stat = "identity")
```



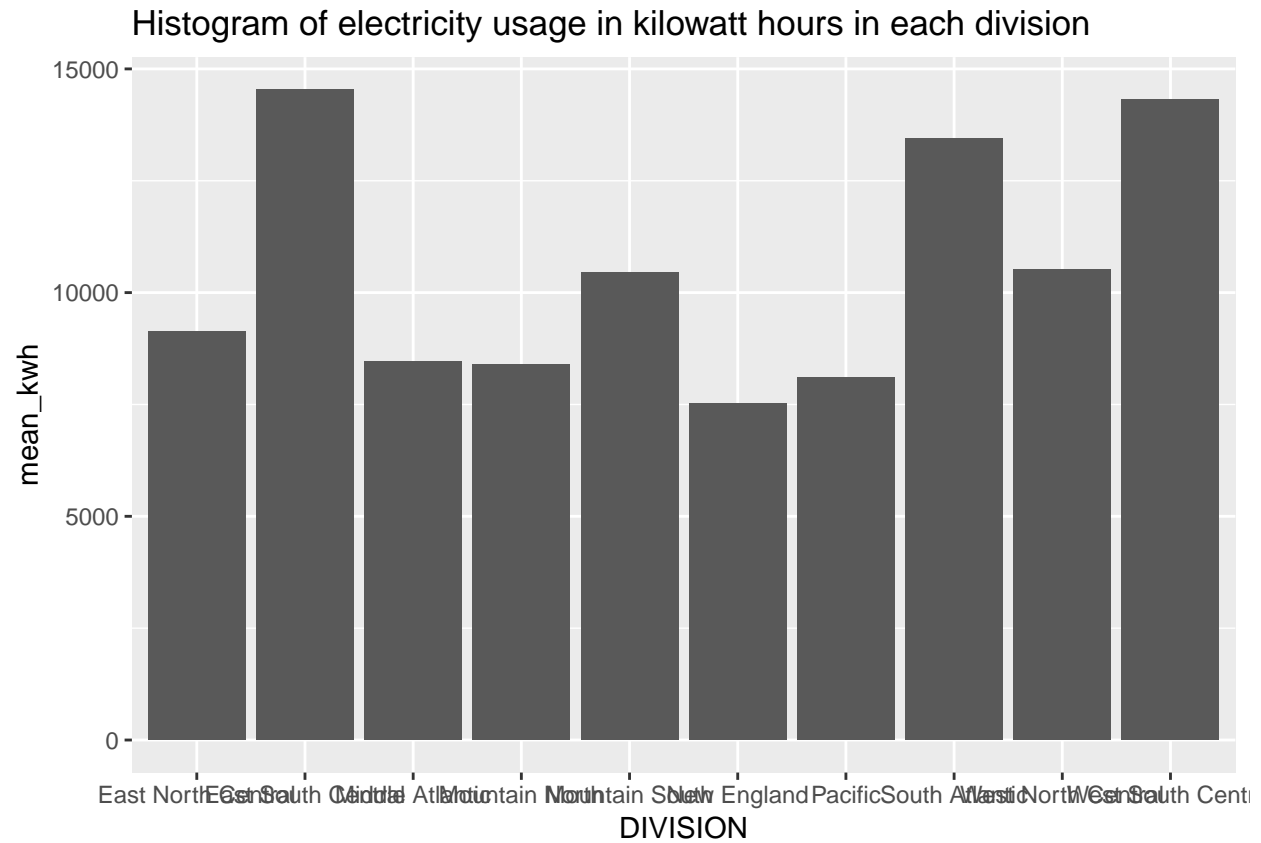
b.

The chart shows average total electricity usage in kilowatt hours in each division.

```
knitr::kable(kwh_devison_sum)
```

DIVISION	mean_kwh
New England	7514.561
Middle Atlantic	8465.442
East North Central	9128.663
West North Central	10523.828
South Atlantic	13446.621
East South Central	14535.969
West South Central	14324.259
Mountain North	8384.471
Mountain South	10442.017
Pacific	8100.405

```
ggplot(kwh_devison_sum,aes(x = DIVISION, y = mean_kwh)) + geom_bar(stat = "identity") + ggtitle("Histogram of Average Total Electricity Usage by Division")
```

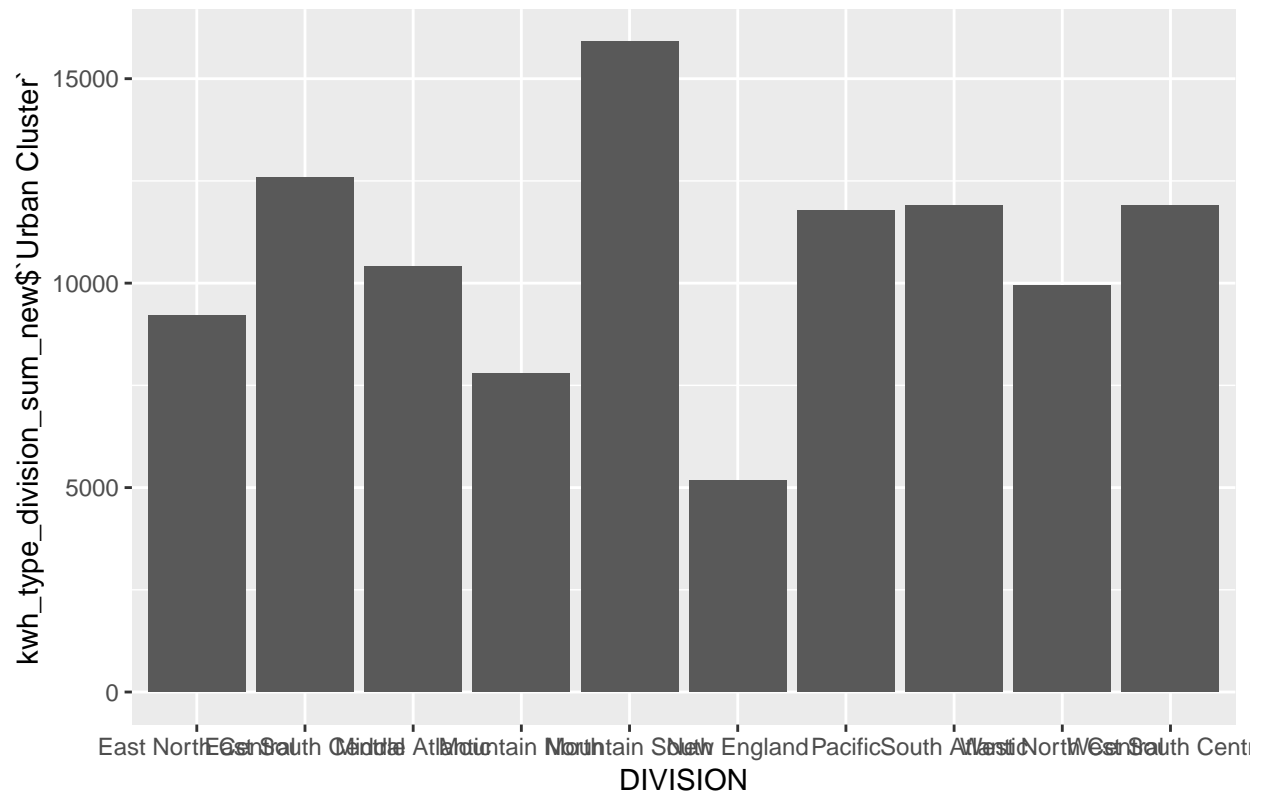


```
knitr::kable(kwh_type_division_sum_new)
```

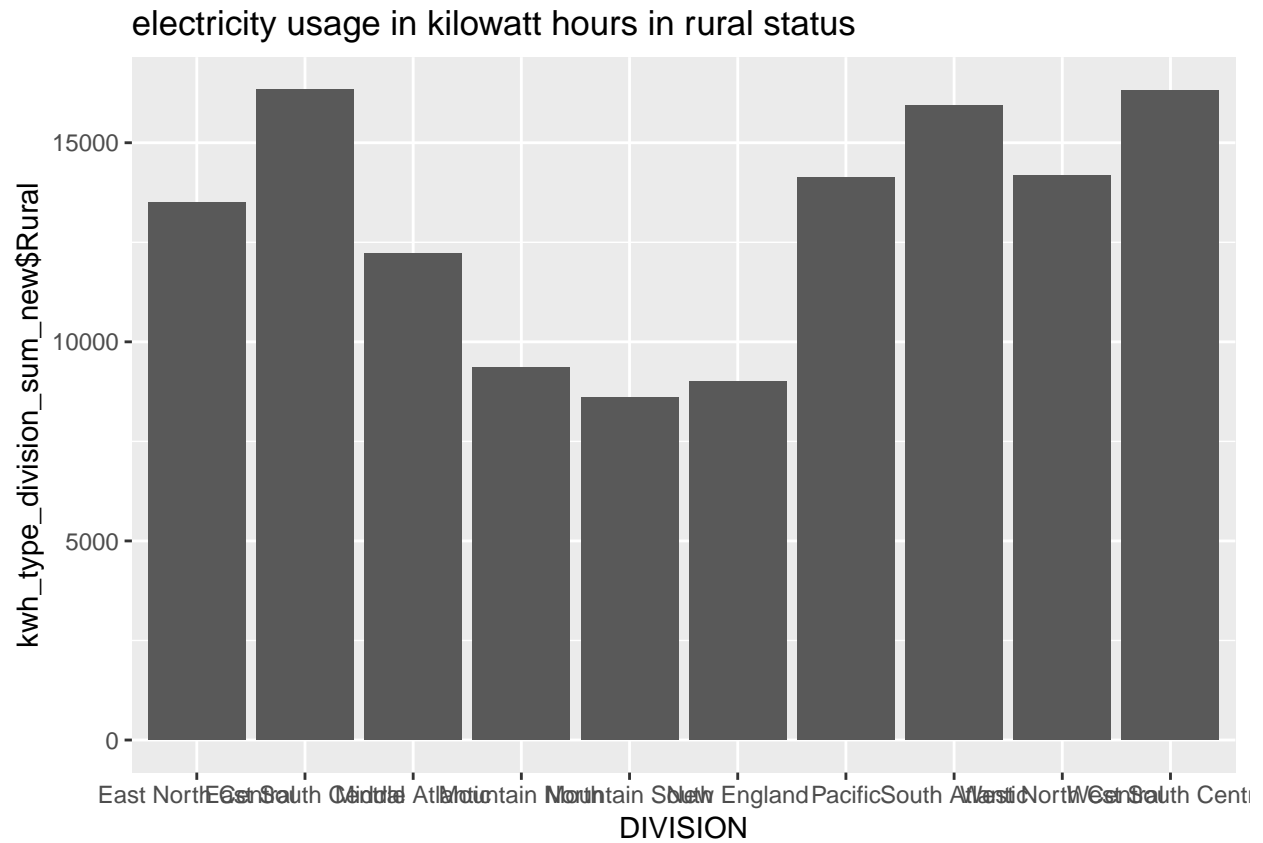
DIVISION	Urban Cluster	Rural	Urban Area
New England	5182.408	9001.054	7626.553
Middle Atlantic	10398.891	12223.389	7788.251
East North Central	9219.739	13500.024	7775.891
West North Central	9936.639	14173.932	9320.237
South Atlantic	11898.968	15941.991	12825.487
East South Central	12590.762	16332.679	14168.218
West South Central	11889.259	16317.160	14061.286
Mountain North	7782.808	9356.001	8143.720
Mountain South	15905.674	8610.426	10670.228
Pacific	11774.131	14114.760	7049.627

```
ggplot(kwh_type_division_sum_new, aes(x = DIVISION, y = kwh_type_division_sum_new$`Urban Cluster`)) + geom_bar()
```

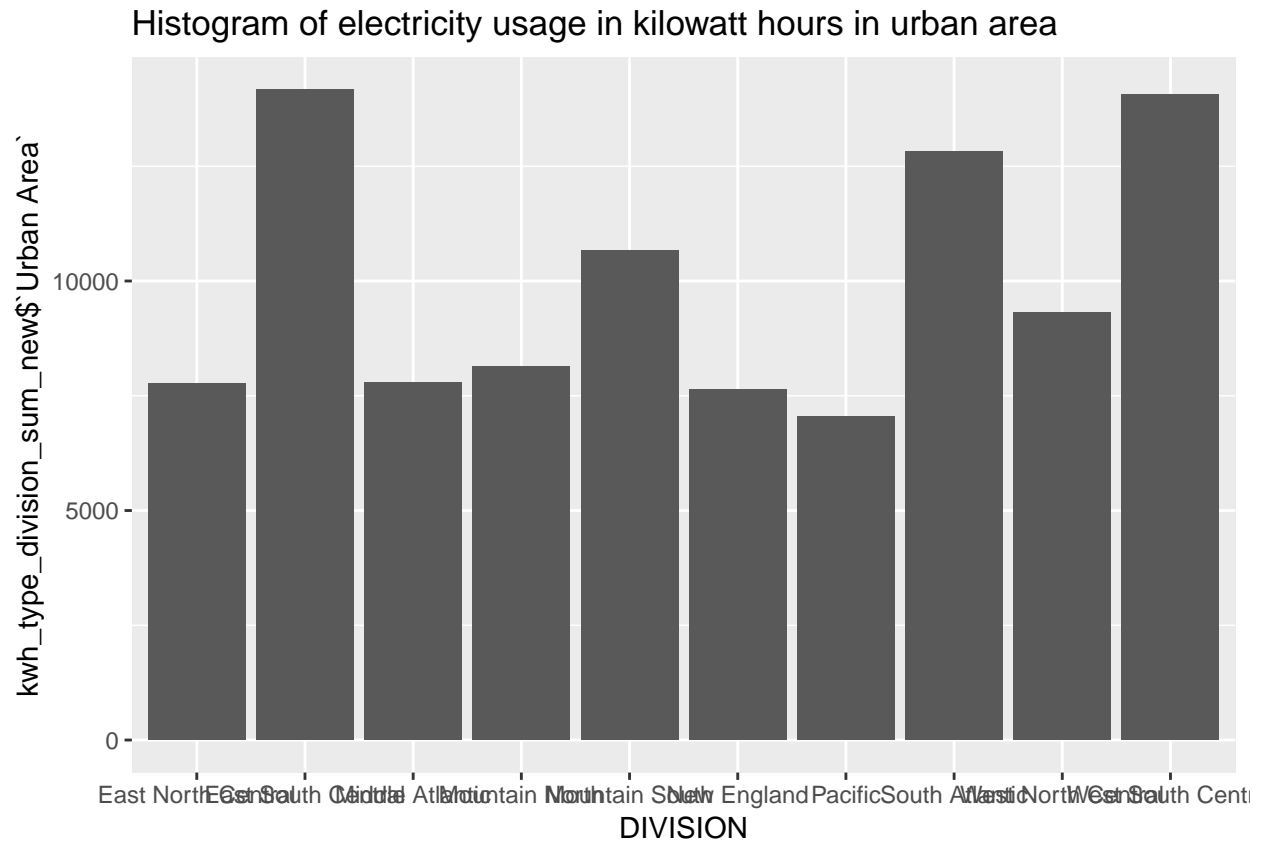
Histogram of electricity usage in kilowatt hours in each divisionif urban cl



```
ggplot(kwh_type_division_sum_new, aes(x = DIVISION, y = kwh_type_division_sum_new$Rural)) + geom_bar(stat = "sum")
```



```
ggplot(kwh_type_division_sum_new, aes(x = DIVISION, y = kwh_type_division_sum_new$Urban Area)) + geom_bar()
```



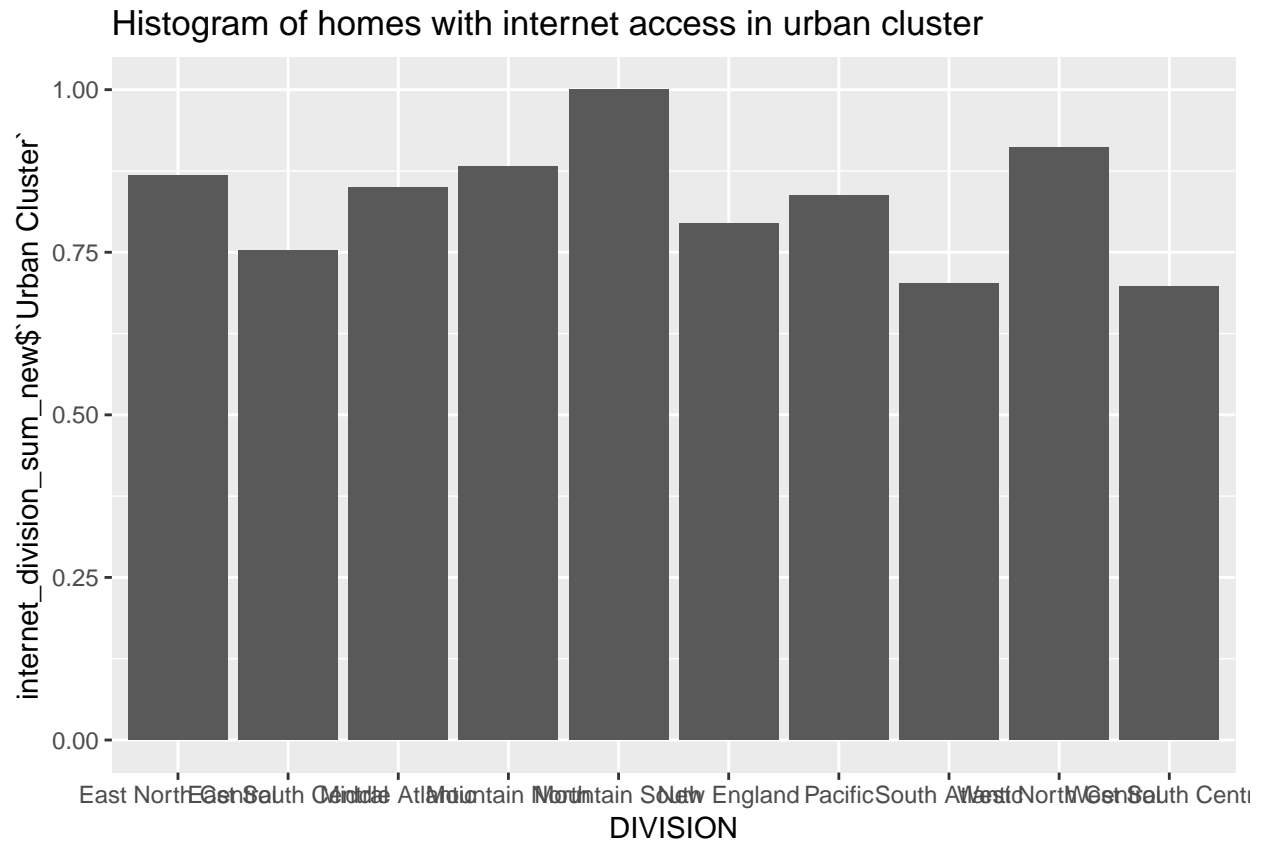
c.

The chart shows the division that has the largest disparity between urban and rural areas in terms of the proportion of homes with internet access.

```
knitr::kable(internet_division_sum_new)
```

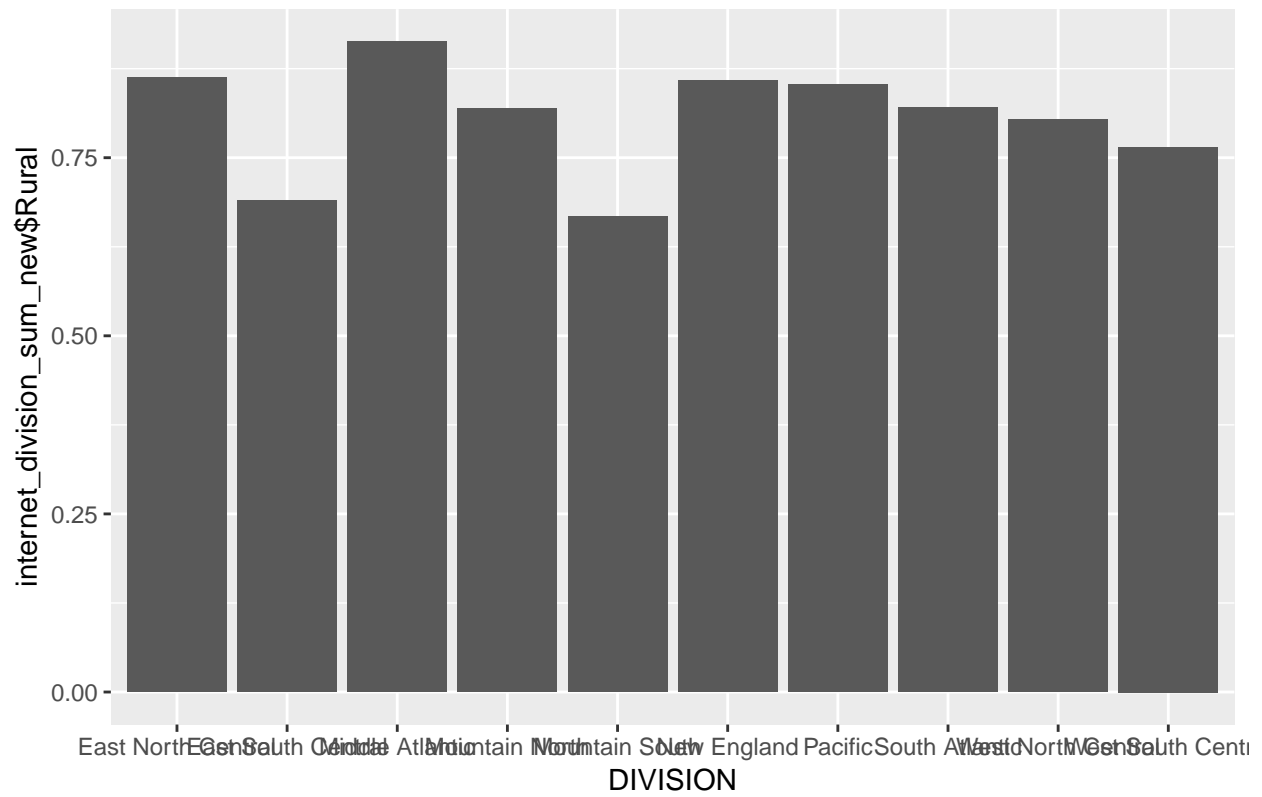
DIVISION	Urban Cluster	Rural	Urban Area
New England	0.7943140	0.8579149	0.9059971
Middle Atlantic	0.8495432	0.9128762	0.8969958
East North Central	0.8685601	0.8621029	0.8615229
West North Central	0.9111256	0.8032515	0.8703800
South Atlantic	0.7025356	0.8204097	0.8713557
East South Central	0.7525202	0.6902596	0.7949211
West South Central	0.6972465	0.7650275	0.8456120
Mountain North	0.8819129	0.8192627	0.8731440
Mountain South	1.0000000	0.6674906	0.8505862
Pacific	0.8368530	0.8527859	0.8905076

```
ggplot(internet_division_sum_new,aes(x = DIVISION, y = internet_division_sum_new$`Urban Cluster`)) + geom_bar()
```

```
ggplot(internet_division_sum_new, aes(x = DIVISION, y = internet_division_sum_new$Rural)) + geom_bar(
```

Histogram of homes with internet access in rural stutas



```
ggplot(internet_division_sum_new, aes(x = DIVISION, y = internet_division_sum_new$`Urban Area`)) + geom_bar()
```

Histogram of homes with internet access in urban area

