Problem Set 1

Israel Diego

October 30, 2018

Question 1

Using command line tools with 2015 Residential Energy Consumption Survey (RECS 2015) data set

Part A

i. This command counts the total number of rows for Region 3

```
cut -d ',' -f 2 recs2015_public_v3.csv | grep 3 | wc -l
```

ii. This command takes the column variables: DOEID, NWEIGHT, andBRRWT1-BRRWT96 and saves the data into a compressed file named 'newData'

```
cut -d ',' -f 1,475-571 recs2015_public_v3.csv | gzip > newData.gz
```

Part B

i. This is a Bash for loop that counts and prints the number of observations within each region

```
for region in 1 2 3 4
do

  echo $region
  cut -d ',' -f 2 recs2015_public_v3.csv| grep $region | wc -l

done
```

ii. This command produces a file named region_division.txt and provides a sorted list showing unique combinations of values from REGIONC and DIVISION.

```
cut -d ',' -f 2,3 recs2015_public_v3.csv| sort -n | uniq > region_division.txt
```

Question 2:

Part A

In this problem we show which airlines were responsible for at least 1% of the flights departing any of the three NYC airports between January 1 and October 31, 2013. The results are shown in the following tables.

Table 1: **NYC Flights Jan1 - Oct31, 2013** Proportions per Airline representing more than 1% of total flights in NYC.

Airline	Flights Per Airline	Proportion of Total Flights (%)
AirTran Airways Corporation	2845	1.01
American Airlines Inc.	27447	9.75
Delta Air Lines Inc.	40168	14.30
Endeavor Air Inc.	15232	5.41
Envoy Air	22202	7.89
ExpressJet Airlines Inc.	45395	16.10
JetBlue Airways	45605	16.20
Southwest Airlines Co.	10143	3.60
United Air Lines Inc.	48880	17.40
US Airways Inc.	17232	6.12
Virgin America	4235	1.51

Part B

Here we compare the number and percent of annual flights using the airlines from Part A, but in year 2014 instead.

Table 2: **2013 Flights:** We start by summarizing the same data from Part A, but now we include CIs for each point estimate.

Airline	# Flights 2013	(%) 2013	95% CI
AirTran Airways Corporation	2845	1.01	[0.974, 1.05]
American Airlines Inc.	27447	9.75	[9.65, 9.86]
Delta Air Lines Inc.	40168	14.30	[14.1, 14.4]
Envoy Air	22202	7.89	[7.79, 7.99]
ExpressJet Airlines Inc.	45395	16.10	[16, 16.3]
JetBlue Airways	45605	16.20	[16.1, 16.3]
Southwest Airlines Co.	10143	3.60	[3.54, 3.67]
United Air Lines Inc.	48880	17.40	[17.2, 17.5]
US Airways Inc.	17232	6.12	[6.04, 6.21]
Virgin America	4235	1.51	[1.46, 1.55]

Table 3: **2014 Flights**

Airline	# Flights 2014	(%) 2014	95% CI
AirTran Airways Corporation	1251	0.494	[0.467, 0.521]
American Airlines Inc.	26302	10.400	[10.3, 10.5]
Delta Air Lines Inc.	41683	16.500	[16.3, 16.6]
Envoy Air	18559	7.330	[7.22, 7.43]
ExpressJet Airlines Inc.	39819	15.700	[15.6, 15.9]

Airline	# Flights 2014	(%) 2014	95% CI
JetBlue Airways	44479	17.600	[17.4, 17.7]
Southwest Airlines Co.	11902	4.700	[4.62, 4.78]
United Air Lines Inc.	46267	18.300	[18.1, 18.4]
US Airways Inc.	16750	6.610	[6.52, 6.71]
Virgin America	4797	1.890	[1.84, 1.95]

Table 4: **Percent Changes from 2013 to 2014**

Airline	(%) Change	CI on (%) Change
AirTran Airways Corporation	-0.517	[-0.563, -0.471]
American Airlines Inc.	0.628	[0.467, 0.79]
Delta Air Lines Inc.	2.180	[1.99, 2.37]
Envoy Air	-0.564	[-0.706, -0.422]
ExpressJet Airlines Inc.	-0.414	[-0.611, -0.218]
JetBlue Airways	1.350	[1.15, 1.55]
Southwest Airlines Co.	1.090	[0.986, 1.2]
United Air Lines Inc.	0.893	[0.687, 1.1]
US Airways Inc.	0.488	[0.357, 0.619]
Virgin America	0.389	[0.319, 0.458]

The Flights with the Largest Percent Increase and Decrease from 2013 to 2014 are shown on the table below.

Table 5: **Airlines with Largest Increase and Decrease:** First Row shows Largest Increase. Second row shows Largest Decrease. Rows are sorted alphabetically.

Airline	(%) Change
Delta Air Lines Inc.	2.180
Envoy Air	-0.564

We notice that some airlines show an increase in the percent of flights, even though their number of flights decreased. Take United Airlines Inc for example. They had 48,880 flights in 2013 and 46,267 flights in 2014. The main reason being that 2014 had less total flights than the total flights in 2013.

Part C

Now we show the percent flights of all three airports for each airline from Part A. Each table summary of the airports is followed by another table showing the carrier at that airport with the largest percent of flights.

Table 6: **Newark Liberty International Airport (EWR)** This table shows the percent flights the airline represented for each airport and we provide Confidence Intervals of the percent estimates.

Airline	(%)	95% CI
AirTran Airways Corporation	0.00	[0, 0]
American Airlines Inc.	2.89	[2.79, 2.98]
Delta Air Lines Inc.	3.59	[3.49, 3.7]

Airline	(%)	95% CI
Endeavor Air Inc.	1.05	[0.992, 1.11]
Envoy Air	1.88	[1.81, 1.96]
ExpressJet Airlines Inc.	36.40	[36.1, 36.6]
JetBlue Airways	5.43	[5.3, 5.55]
Southwest Airlines Co.	5.12	[5, 5.25]
United Air Lines Inc.	38.10	[37.9, 38.4]
US Airways Inc.	3.65	[3.54, 3.75]
Virgin America	1.30	[1.23, 1.36]

Table 7: **Largest Carrier at Newark Liberty International Airport (EWR)**

Airline	(%)	95% CI
United Air Lines Inc.	38.1	[37.9, 38.4]

Table 8: **John F. Kennedy International Airport (JFK)** This table shows the percent flights the airline represented for each airport and we provide Confidence Intervals of the percent estimates.

Airline	(%)	95% CI
AirTran Airways Corporation	0.00	[0, 0]
American Airlines Inc.	12.40	[12.2, 12.6]
Delta Air Lines Inc.	18.60	[18.4, 18.8]
Endeavor Air Inc.	13.20	[13, 13.4]
Envoy Air	6.46	[6.32, 6.61]
ExpressJet Airlines Inc.	1.27	[1.2, 1.33]
JetBlue Airways	37.80	[37.5, 38.1]
Southwest Airlines Co.	0.00	[0, 0]
United Air Lines Inc.	4.07	[3.96, 4.19]
US Airways Inc.	2.69	[2.6, 2.79]
Virgin America	3.23	[3.13, 3.34]

Table 9: **Largest Carrier at John F. Kennedy International Airport (JFK)**

Airline	(%)	95% CI
JetBlue Airways	37.8	[37.5, 38.1]

Table 10: **LaGuardia Airport (LGA)** This table shows the percent flights the airline represented for each airport and we provide Confidence Intervals of the percent estimates.

Airline	(%)	95% CI
AirTran Airways Corporation	3.11	[3.01, 3.22]
American Airlines Inc.	14.80	[14.6, 15]
Delta Air Lines Inc.	22.00	[21.8, 22.3]

Airline	(%)	95% CI
Endeavor Air Inc.	2.43	[2.33, 2.52]
Envoy Air	16.20	[16, 16.4]
ExpressJet Airlines Inc.	8.43	[8.26, 8.6]
JetBlue Airways	5.73	[5.59, 5.88]
Southwest Airlines Co.	5.82	[5.67, 5.96]
United Air Lines Inc.	7.69	[7.52, 7.85]
US Airways Inc.	12.60	[12.4, 12.8]
Virgin America	0.00	[0, 0]

Table 11: **Largest Carrier at LaGuardia Airport (LGA)**

Airline	(%)	95% CI
Delta Air Lines Inc.	22	[21.8, 22.3]

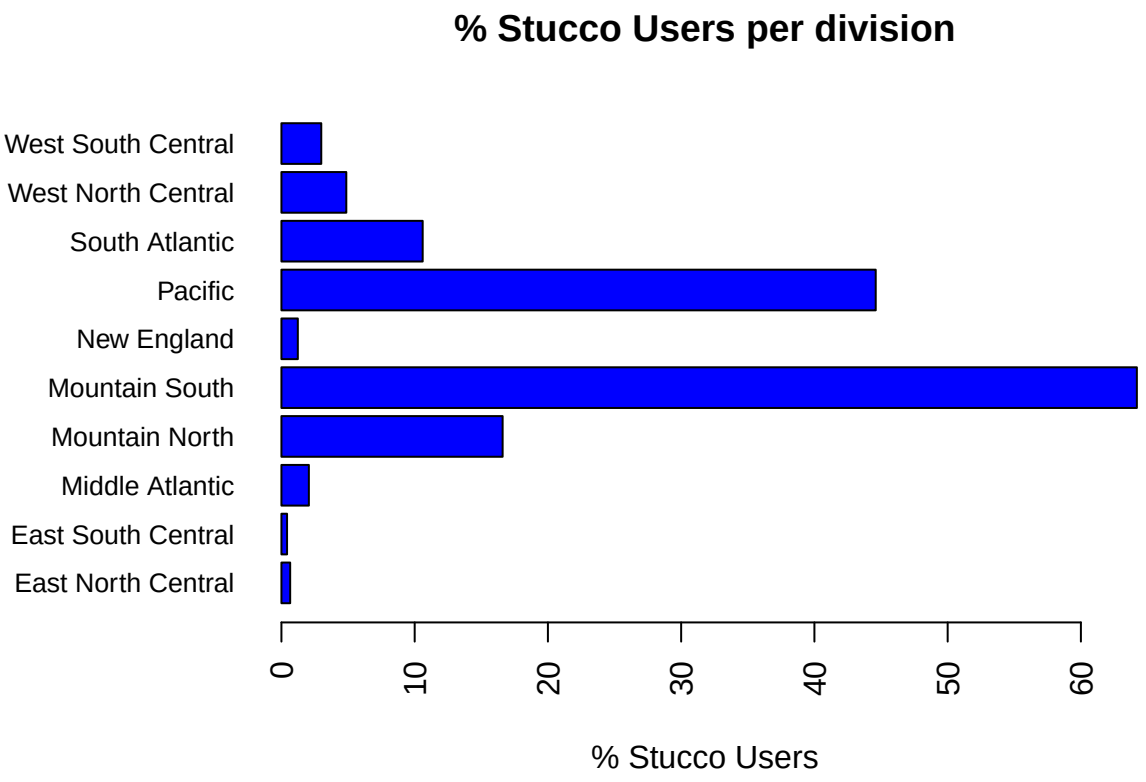
Question 3:

Part A

Using the RECS 2015 data, first we compute the percent of homes that have stucco construction as their major outside wall material within each division. We compute (RSEs) using the procedure provided in the RECS 2015 documentation.

Table 12: **Percent of Stucco Wall Users per Division** For each division we include the percentage of Homes in that division that have Stucco construction as main wall material. We also include Standard Errors and Confidence Intervals of each estimate.

Division	(%) Stucco	Std Error	95% CI
East North Central	0.657	0.283	[0.101, 1.21]
East South Central	0.423	0.409	[-0.379, 1.22]
Middle Atlantic	2.060	0.734	[0.622, 3.5]
Mountain North	16.600	3.240	[10.2, 23]
Mountain South	64.200	4.490	[55.4, 73]
New England	1.230	0.801	[-0.343, 2.8]
Pacific	44.600	1.690	[41.3, 47.9]
South Atlantic	10.600	1.420	[7.83, 13.4]
West North Central	4.870	2.030	[0.891, 8.84]
West South Central	2.990	0.694	[1.63, 4.35]

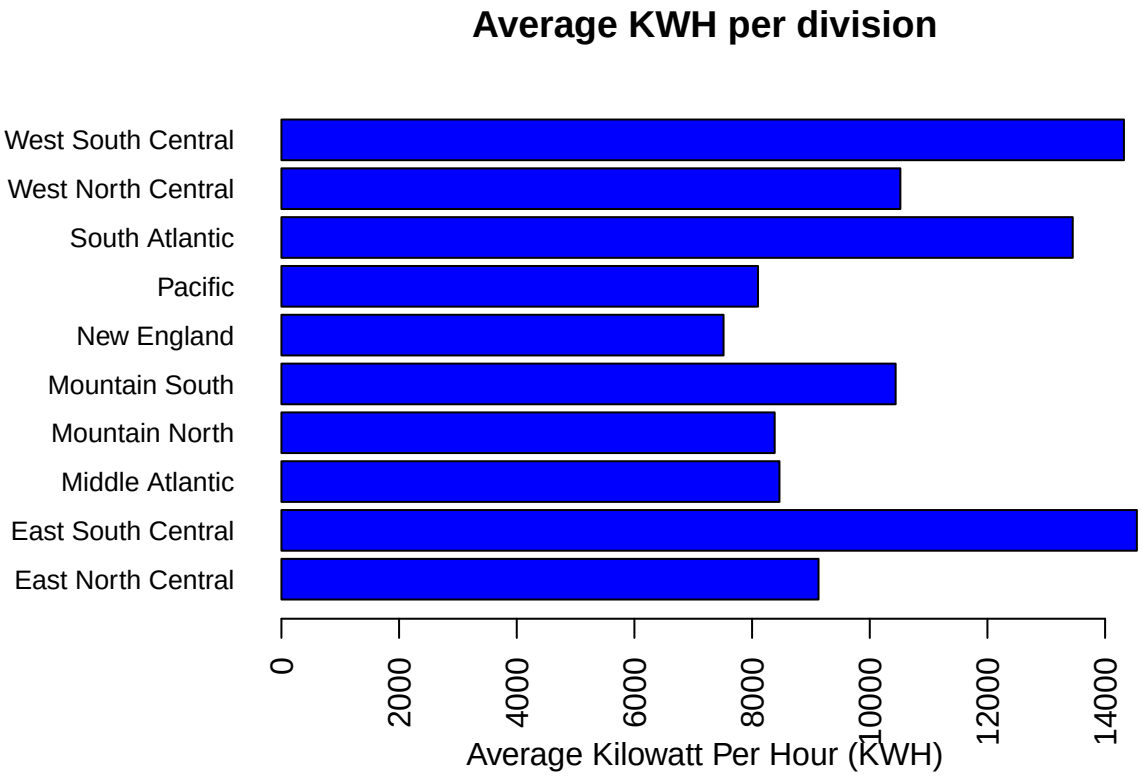


Part B

Now we show average total electricity usage in kilowatt hours per division. We illustrate our results in the table below and also graphically with a bar chart.

Table 13: **Average Total Electricity Usage in Kilowatt hours Per Division** For each division we include the Average total KWH used. We also include Standard Errors and Confidence Intervals of each estimate.

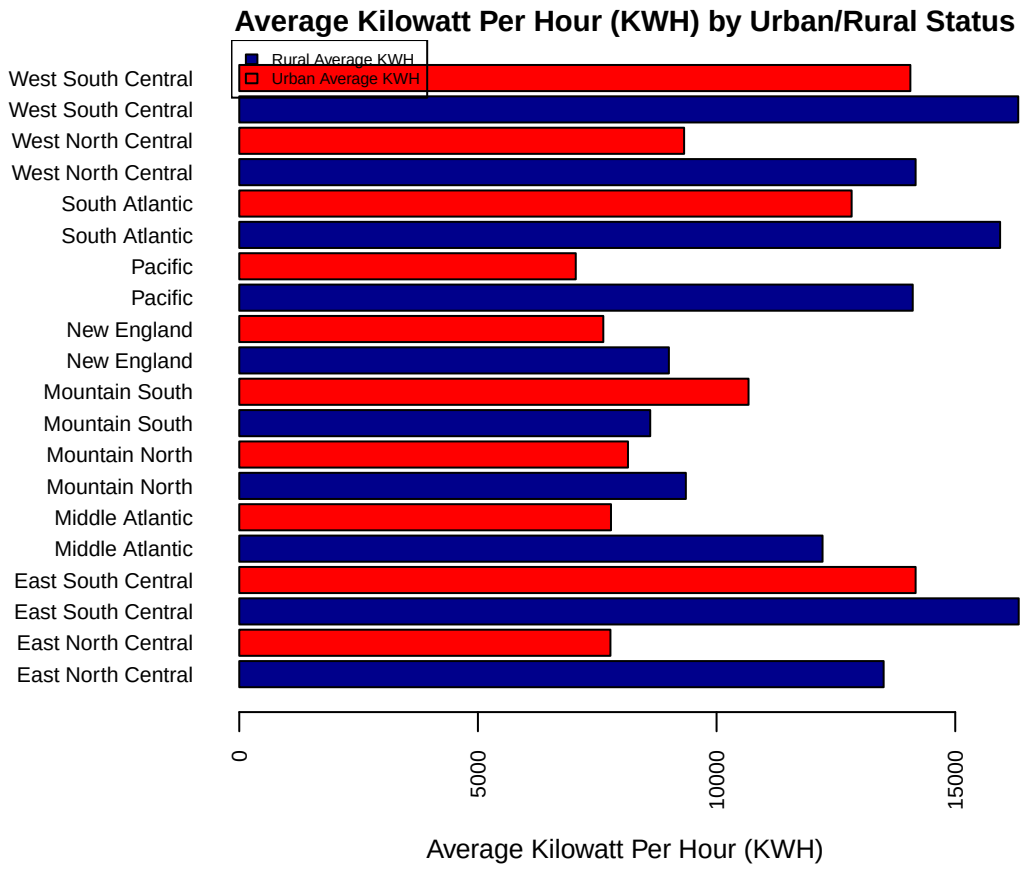
Division	Average KWH	Std Error	95% CI
East North Central	9129	203.6	[8730, 9528]
East South Central	14540	620.6	[13320, 15750]
Middle Atlantic	8465	201.1	[8071, 8860]
Mountain North	8384	644.4	[7121, 9648]
Mountain South	10440	1271.0	[7950, 12930]
New England	7515	532.0	[6472, 8557]
Pacific	8100	178.6	[7750, 8450]
South Atlantic	13450	276.9	[12900, 13990]
West North Central	10520	453.6	[9635, 11410]
West South Central	14320	423.0	[13500, 15150]



From the bar chart above, East South Central has the highest average total electricity usage. Next, we would like to compute average total electricity usage stratified by urban and rural status.

Table 14: **Average Total Electricity Usage in Kilowatt hours Per Division by Urban and Rural Status:** For each division we include the Average total KWH used in the Urban and Rural categories. We include Standard Errors and Confidence Intervals of each estimate.

Division	Urban/Rural	Average KWH	Std Error	95% CI
East North Central	Rural	13500	754.3	[12020, 14980]
East North Central	Urban	7776	254.6	[7277, 8275]
East South Central	Rural	16330	1146.0	[14090, 18580]
East South Central	Urban	14170	1039.0	[12130, 16200]
Middle Atlantic	Rural	12220	811.4	[10630, 13810]
Middle Atlantic	Urban	7788	214.8	[7367, 8209]
Mountain North	Rural	9356	1866.0	[5698, 13010]
Mountain North	Urban	8144	361.4	[7435, 8852]
Mountain South	Rural	8610	1058.0	[6536, 10680]
Mountain South	Urban	10670	1305.0	[8112, 13230]
New England	Rural	9001	1140.0	[6766, 11240]
New England	Urban	7627	549.6	[6549, 8704]
Pacific	Rural	14110	1079.0	[12000, 16230]
Pacific	Urban	7050	314.4	[6433, 7666]
South Atlantic	Rural	15940	563.0	[14840, 17050]
South Atlantic	Urban	12830	361.4	[12120, 13530]
West North Central	Rural	14170	798.8	[12610, 15740]
West North Central	Urban	9320	540.3	[8261, 10380]
West South Central	Rural	16320	1148.0	[14070, 18570]
West South Central	Urban	14060	459.5	[13160, 14960]



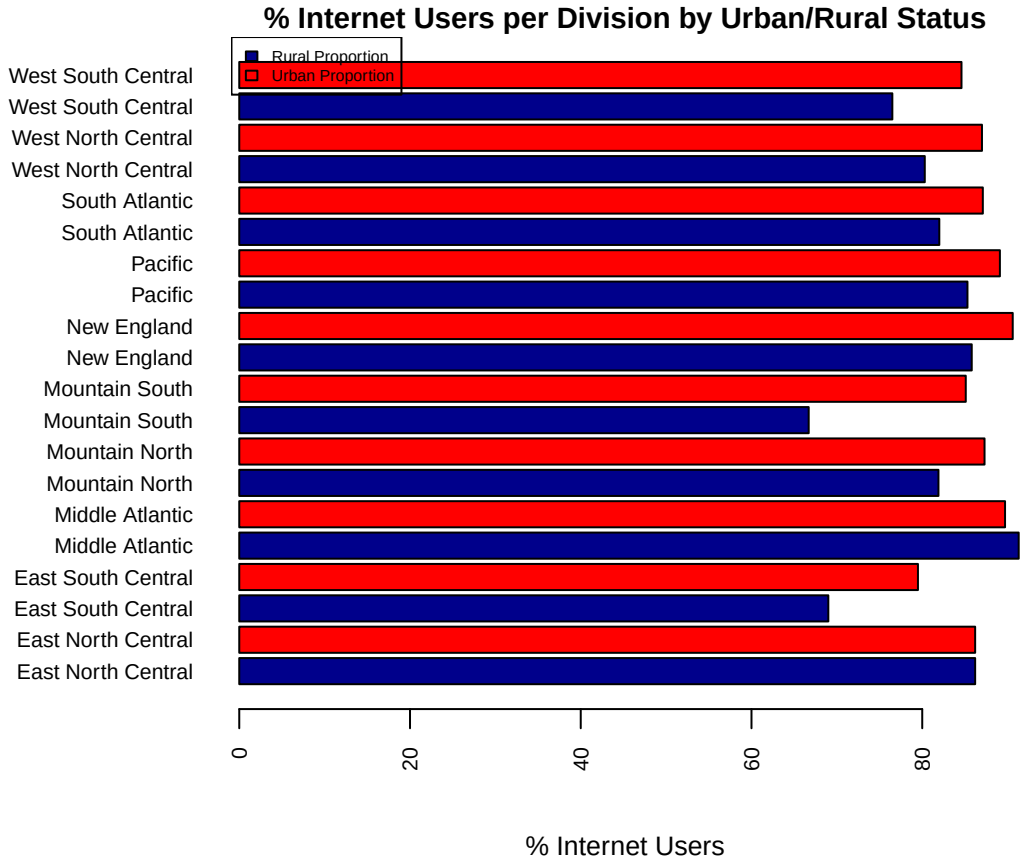
Part C

Here we show percents of Internet Users by Division and by Urban and Rural Status. We illustrate our results in the table below and also graphically with a bar chart. The goal is to identify the division with the largest disparity between Urban and Rural percentages of Internet Users.

Table 15: **Percentage of Internet Users by Division and by Urban and Rural Status:** For each division we include the percent of homes that use internet in the Urban and Rural categories. We include Standard Errors and Confidence Intervals of each estimate.

Division	Urban/Rural	(%) Internet	Std Error	95% CI
East North Central	Rural	86.2	2.33	[81.6, 90.8]
East North Central	Urban	86.2	1.51	[83.2, 89.1]
East South Central	Rural	69.0	2.82	[63.5, 74.6]
East South Central	Urban	79.5	5.52	[68.7, 90.3]
Middle Atlantic	Rural	91.3	3.05	[85.3, 97.3]
Middle Atlantic	Urban	89.7	2.86	[84.1, 95.3]
Mountain North	Rural	81.9	4.14	[73.8, 90]
Mountain North	Urban	87.3	2.77	[81.9, 92.7]
Mountain South	Rural	66.7	4.33	[58.3, 75.2]
Mountain South	Urban	85.1	2.15	[80.8, 89.3]
New England	Rural	85.8	1.75	[82.4, 89.2]
New England	Urban	90.6	1.90	[86.9, 94.3]

Division	Urban/Rural	(%) Internet	Std Error	95% CI
Pacific	Rural	85.3	4.00	[77.4, 93.1]
Pacific	Urban	89.1	1.36	[86.4, 91.7]
South Atlantic	Rural	82.0	2.94	[76.3, 87.8]
South Atlantic	Urban	87.1	1.64	[83.9, 90.4]
West North Central	Rural	80.3	4.51	[71.5, 89.2]
West North Central	Urban	87.0	2.02	[83.1, 91]
West South Central	Rural	76.5	2.23	[72.1, 80.9]
West South Central	Urban	84.6	3.06	[78.6, 90.6]



In the table below, we show the division with the largest absolute difference between Urban and Rural percentage. Mountain South has the largest disparity between percent of Urban and Rural Internet Users.

Table 16: **Division with Largest Disparity between Urban and Rural Percent of Internet users**

Division	(%) Rural	(%) Urban	Abs Diff
Mountain South	66.7	85.1	18.4