Stat506_PS1 Part 2 & 3 Meng-Ni Ho

Question 2a

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?

```
## # A tibble: 11 x 6
##    carrier n_2013 p_2013 se_2013 CI_lwr_2013 CI_upr_2013
##    <chr>    <int>  <dbl>   <dbl>       <dbl>       <dbl>
##  1 9E       15232  0.054   0.002       0.05        0.058
##  2 AA       27447  0.098   0.002       0.094       0.102
##  3 B6       45605  0.162   0.002       0.158       0.166
##  4 DL       40168  0.143   0.002       0.139       0.147
##  5 EV       45395  0.161   0.002       0.157       0.165
##  6 FL        2845  0.01    0.002       0.006       0.014
##  7 MQ       22202  0.079   0.002       0.075       0.083
##  8 UA       48880  0.174   0.002       0.17        0.178
##  9 US       17232  0.061   0.002       0.057       0.065
## 10 VX        4235  0.015   0.002       0.011       0.019
## 11 WN       10143  0.036   0.002       0.032       0.04
```

Question 2b

Among the airlines from part "a", compare the number and percent of annual flights in the first 10 months of 2013 and the first 10 months of 2014.

```
##    carrier n_2013 p_2013 se_2013 CI_lwr_2013 CI_upr_2013 n_2014 p_2014
## 1       AA  27447  0.098   0.002       0.094       0.102  26302  0.104
## 2       B6  45605  0.162   0.002       0.158       0.166  44479  0.176
## 3       DL  40168  0.143   0.002       0.139       0.147  41683  0.165
## 4       EV  45395  0.161   0.002       0.157       0.165  39819  0.157
## 5       FL   2845  0.010   0.002       0.006       0.014   1251  0.005
## 6       MQ  22202  0.079   0.002       0.075       0.083  18559  0.073
## 7       UA  48880  0.174   0.002       0.170       0.178  46267  0.183
## 8       US  17232  0.061   0.002       0.057       0.065  16750  0.066
## 9       VX   4235  0.015   0.002       0.011       0.019   4797  0.019
## 10      WN  10143  0.036   0.002       0.032       0.040  11902  0.047
##    se_2014 CI_lwr_2014 CI_upr_2014                      name p_change
## 1    0.002       0.100       0.108      American Airlines Inc.    0.006
## 2    0.002       0.172       0.180             JetBlue Airways    0.014
## 3    0.002       0.161       0.169        Delta Air Lines Inc.    0.022
## 4    0.002       0.153       0.161    ExpressJet Airlines Inc.   -0.004
## 5    0.002       0.001       0.009 AirTran Airways Corporation   -0.005
## 6    0.002       0.069       0.077                   Envoy Air   -0.006
## 7    0.002       0.179       0.187       United Air Lines Inc.    0.009
## 8    0.002       0.062       0.070             US Airways Inc.    0.005
## 9    0.002       0.015       0.023              Virgin America    0.004
## 10   0.002       0.043       0.051      Southwest Airlines Co.    0.011
##    n_change
## 1     -1145
## 2     -1126
```

```
## 3      1515
## 4     -5576
## 5     -1594
## 6     -3643
## 7     -2613
## 8      -482
## 9       562
## 10     1759
```

1. Which airline showed the largest increase?

```
sqldf("select carrier, name, max(p_change) from merge1")
```

```
##   carrier                 name max(p_change)
## 1      DL Delta Air Lines Inc.         0.022
```

2. Which airline showed the largest decrease?

```
sqldf("select carrier, name, min(p_change) from merge1")
```

```
##   carrier      name min(p_change)
## 1      MQ Envoy Air        -0.006
```

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

Becasue total flights in 2014 decreased compared to 2013, then for the airlines that showed decreased flights, their percentage will increase.

Question 2c

Among of the three NYC airports, produce a table showing the percent of flights each airline is responsible for. Limit the table to the airlines identified in part a and include confidence intervals for your estimates.

```
##    origin carrier n_flights  total proportion    se CI_lwr CI_upr
## 1     EWR      9E      1268 120115      0.011 0.003  0.005  0.017
## 2     EWR      AA      3487 120115      0.029 0.003  0.023  0.035
## 3     EWR      B6      6557 120115      0.055 0.003  0.049  0.061
## 4     EWR      DL      4342 120115      0.036 0.003  0.030  0.042
## 5     EWR      EV     43939 120115      0.366 0.002  0.362  0.370
## 6     EWR      MQ      2276 120115      0.019 0.003  0.013  0.025
## 7     EWR      UA     46087 120115      0.384 0.002  0.380  0.388
## 8     EWR      US      4405 120115      0.037 0.003  0.031  0.043
## 9     EWR      VX      1566 120115      0.013 0.003  0.007  0.019
## 10    EWR      WN      6188 120115      0.052 0.003  0.046  0.058
## 11    JFK      9E     14651 110937      0.132 0.003  0.126  0.138
## 12    JFK      AA     13783 110937      0.124 0.003  0.118  0.130
## 13    JFK      B6     42076 110937      0.379 0.002  0.375  0.383
## 14    JFK      DL     20701 110937      0.187 0.003  0.181  0.193
## 15    JFK      EV      1408 110937      0.013 0.003  0.007  0.019
## 16    JFK      MQ      7193 110937      0.065 0.003  0.059  0.071
## 17    JFK      UA      4534 110937      0.041 0.003  0.035  0.047
## 18    JFK      US      2995 110937      0.027 0.003  0.021  0.033
## 19    JFK      VX      3596 110937      0.032 0.003  0.026  0.038
## 20    LGA      9E      2541 103350      0.025 0.003  0.019  0.031
## 21    LGA      AA     15459 103350      0.150 0.003  0.144  0.156
## 22    LGA      B6      6002 103350      0.058 0.003  0.052  0.064
## 23    LGA      DL     23067 103350      0.223 0.003  0.217  0.229
```

```
## 24    LGA      EV      8826 103350      0.085 0.003  0.079  0.091
## 25    LGA      FL      3260 103350      0.032 0.003  0.026  0.038
## 26    LGA      MQ     16928 103350      0.164 0.003  0.158  0.170
## 27    LGA      UA      8044 103350      0.078 0.003  0.072  0.084
## 28    LGA      US     13136 103350      0.127 0.003  0.121  0.133
## 29    LGA      WN      6087 103350      0.059 0.003  0.053  0.065
```

- Which airline is the largest carrier at each airport?

```
sqldf("select origin, carrier, max(proportion) as max_ from merge2 group by origin")
```

```
##   origin carrier  max_
## 1    EWR      UA 0.384
## 2    JFK      B6 0.379
## 3    LGA      DL 0.223
```

Question 3a

What percent of homes have stucco construction as the major outside wall material within each division?

```
## # A tibble: 10 x 6
## # Groups:   DIVISION [10]
##    DIVISION WALLTYPE   pct std_err     lwr    upr
##       <int>    <int> <dbl>   <dbl>   <dbl>  <dbl>
##  1        6        4  0.42    1.55   -2.61   1.55
##  2        3        4  0.66    0.89   -1.09   0.89
##  3        1        4  1.23    0.81   -0.36   0.81
##  4        2        4  2.06    2.07   -2      2.07
##  5        7        4  2.99    3.9    -4.65   3.9
##  6        4        4  4.87    7.66  -10.2    7.66
##  7        5        4 10.6    19.2   -26.9   19.2
##  8        8        4 16.6    31.1   -44.4   31.1
##  9       10        4 44.6    87.1  -126.    87.1
## 10        9        4 64.2   126.   -184.   126.
```

1. Which division has the highest proportion?

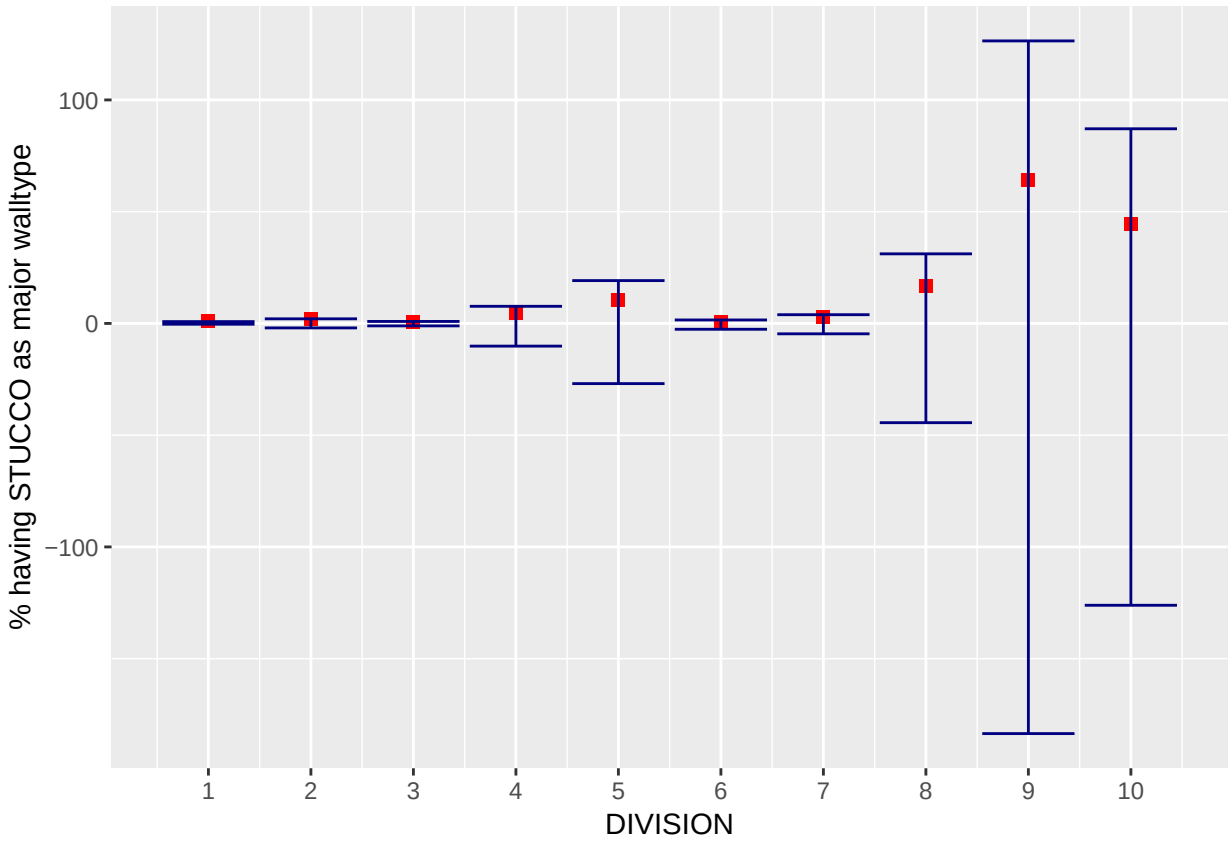
```
sqldf("select DIVISION, max(pct) as max_percentage from stucco_wall_prop")
```

```
##   DIVISION max_percentage
## 1        9          64.25
```

2. Which division has the lowest proportion?

```
sqldf("select DIVISION, min(pct) as min_percentage from stucco_wall_prop")
```

```
##   DIVISION min_percentage
## 1        6           0.42
```

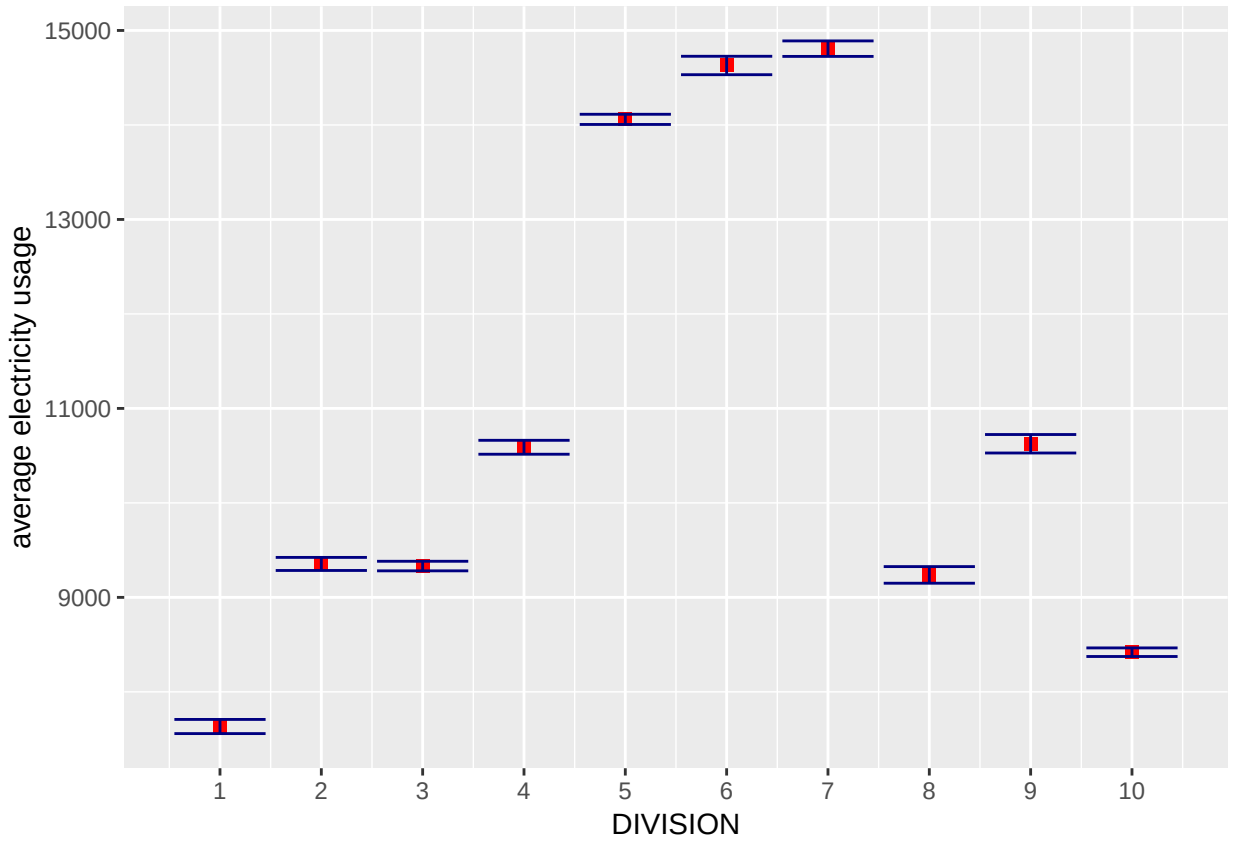


- Plot: % of having STUCCO as major walltype in each division

Question 3b

1. What is average total electricity usage in kilowatt hours in each division?

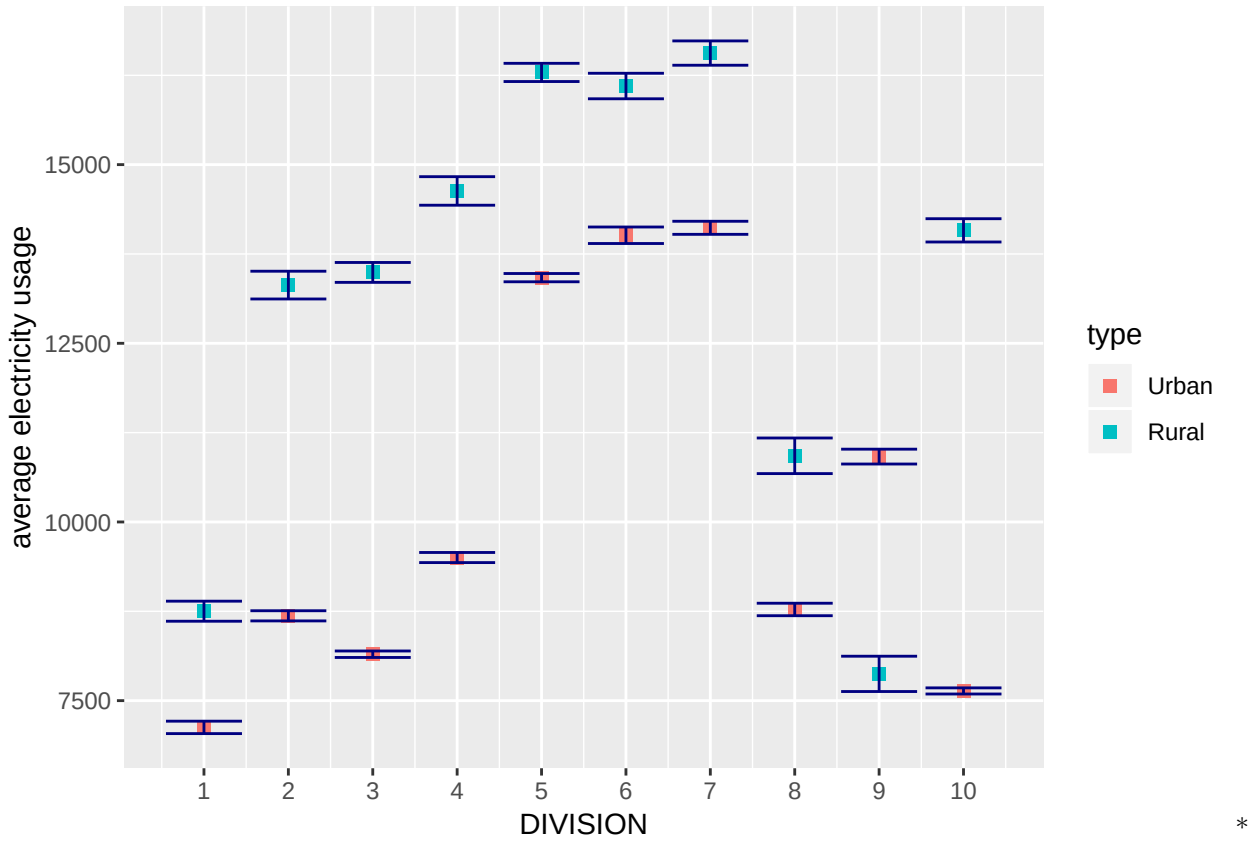
```
## # A tibble: 10 x 7
##    DIVISION mean_elec sd_elec n_elec se_elec    lwr    upr
##       <int>     <dbl>   <dbl>  <int>   <dbl>  <dbl>  <dbl>
##  1        1     7633.   4807.    253    302.  7558.  7708.
##  2        2     9354.   6482.    541    279.  9285.  9423.
##  3        3     9332.   5893.    836    204.  9281.  9382.
##  4        4    10589.   6580.    491    297. 10515. 10662.
##  5        5    14059.   7068.   1058    217. 14005. 14113.
##  6        6    14629.   7588.    372    393. 14532. 14727.
##  7        7    14808.   7967.    580    331. 14725. 14890.
##  8        8     9238.   5354.    228    355.  9150.  9326.
##  9        9    10626.   6132.    242    394. 10528. 10724.
## 10       10     8420.   6029.   1085    183.  8374.  8465.
```



- Plot: Average electricity usage in each division

2. What is average total electricity usage in kilowatt hours in each division stratified by urban and rural?

```
## # A tibble: 20 x 8
## # Groups:   DIVISION [10]
##    DIVISION type  mean_elec sd_elec n_elec se_elec    lwr    upr
##       <int> <fct>     <dbl>   <dbl>  <int>   <dbl>  <dbl>  <dbl>
##  1        1 Urban     7126.   4629.    174    351.  7039.  7213.
##  2        1 Rural     8751.   5028.     79    566.  8611.  8892.
##  3        2 Urban     8687.   6168.    463    287.  8616.  8758.
##  4        2 Rural    13315.   6914.     78    783. 13120. 13509.
##  5        3 Urban     8149.   4665.    651    183.  8104.  8195.
##  6        3 Rural    13492.   7639.    185    562. 13353. 13631.
##  7        4 Urban     9502.   5602.    387    285.  9432.  9573.
##  8        4 Rural    14632.   8210.    104    805. 14432. 14832.
##  9        5 Urban    13419.   6685.    822    233. 13361. 13476.
## 10        5 Rural    16290.   7885.    236    513. 16163. 16418.
## 11        6 Urban    14012.   7536.    262    466. 13897. 14127.
## 12        6 Rural    16100.   7545.    110    719. 15922. 16278.
## 13        7 Urban    14116.   7519.    416    369. 14025. 14208.
## 14        7 Rural    16561.   8788.    164    686. 16391. 16731.
## 15        8 Urban     8776.   4712.    179    352.  8688.  8863
## 16        8 Rural    10926    7034.     49   1005. 10677. 11175.
## 17        9 Urban    10915.   6196.    219    419. 10811  11019.
## 18        9 Rural     7875.   4774.     23    995.  7628.  8121.
## 19       10 Urban     7636.   5340.    953    173.  7593.  7679.
## 20       10 Rural    14081.   7539.    132    656. 13918. 14243.
```



Plot: average total electricity usage in kilowatt hours in each division stratified by urban and rural

Question 3c

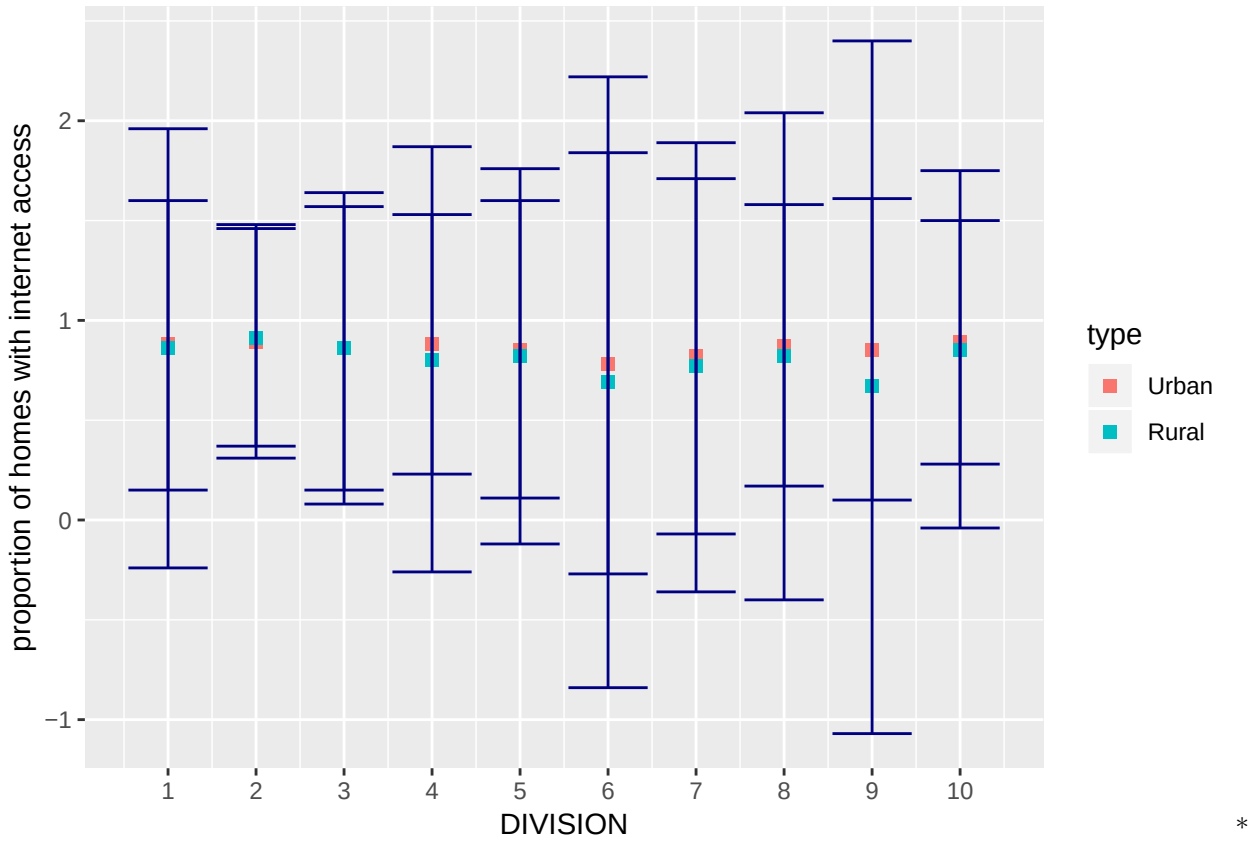
Which division has the largest disparity between urban and rural areas in terms of proportion of homes with internet access?

- Table showing disparity (urban vs rural) in terms of the proportion of homes with internet access?

```
## # A tibble: 10 x 4
## # Groups:   DIVISION [10]
##    DIVISION Urban Rural disparity
##       <int> <dbl> <dbl>     <dbl>
##  1        2  0.89  0.91   -0.02
##  2        3  0.86  0.86    0
##  3        1  0.88  0.86    0.02
##  4        5  0.85  0.82    0.03
##  5       10  0.89  0.85    0.04
##  6        7  0.82  0.77    0.0500
##  7        8  0.87  0.82    0.05
##  8        4  0.88  0.8     0.0800
##  9        6  0.78  0.69    0.09
## 10        9  0.85  0.67    0.180
```

- Which division has the largest disparity between urban and rural areas in terms of the proportion of homes with internet access?

```
##   DIVISION max(disparity)
## 1        9           0.18
```



Plot: proportion of homes with internet access between urban and rural in each division

*