Week 4 Activity, Part 2

Group 5: Zhilin He, Yan Chen, Tianshi Wang, Hao He

9/29/2020

Part 2 (ggplot2)

In this part of the activity, you will use the summary data below and attempt to reproduce the plot in each question using ggplot.

Here is some preparatory work to get the summary results (partially) into shape for plotting.

```
# libraries
library(tidyverse)
```

```
## Warning: package 'tibble' was built under R version 3.6.2
```

```
# read in summary results
q2_df = readr::read_delim("./w4_p2_RECS2015_internet.csv", delim = ",")
```

```
# reshape for plotting
plot_data = q2_df %>%
  pivot_longer(cols = starts_with('p_int'), names_prefix = 'p_int',
               names_to = 'var') %>%
  mutate( rurality = stringr::str_replace_all(var, '_.*', ''),
          type = ifelse( grepl('_', var),
                         stringr::str_replace_all(var, '.*_', ''),
                         'est'
                         )
  ) %>%
  pivot_wider(id_cols = c('division', 'rurality'),
              names_from = type,
              values_from = value
  )
```

1. Write code where indicated to match the first figure included in the README.

```
cap = paste("*Proportion of homes with internet access in each",
            "Census Division.*")

# Use ggplot to produce a figure `p1` matching what is shown using the
#  data above.
library(ggplot2)
library(dplyr)
plot_data_1 = plot_data[plot_data$rurality %in% c('rural', 'urban'), ]
p1 = plot_data_1 %>%
  mutate(est = 100*est, lwr = 100*lwr, upr = 100*upr) %>%
  ggplot(aes(x = division, y = est, color = rurality)) +
  geom_point(position = position_dodge(width = 0.5)) +
  geom_errorbar(
                aes(ymin = lwr, ymax = upr),
```

```
                position = position_dodge(width = 0.5)
               ) +
  theme_bw() +
  coord_flip() +
  scale_color_manual(values = c("rural" = "brown", "urban" = "darkblue")) +
  ylab('Proportion of homes with internet access')
print(p1)
```

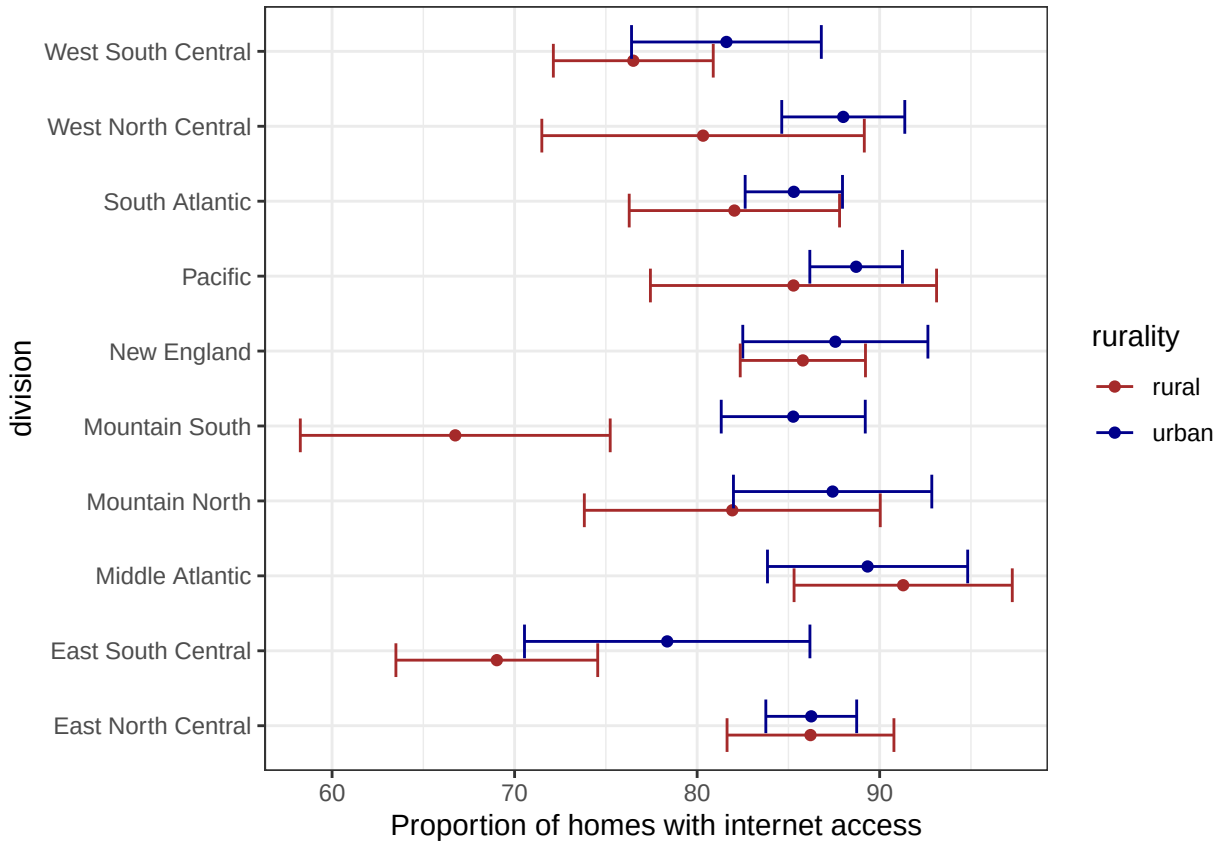

Figure 1: *Proportion of homes with internet access in each Census Division.*

```
# Remove comments to plot and save
ggsave('w4_p2_q1_plot.png', p1)
#p1
```

2. Write code where indicated to match the second figure included in the README.

```
cap = paste("*Difference (urban less rural) in proportion of",
  "homes with internet access in each Census Division.*")

# Use ggplot to produce a figure `p2` matching what is shown using the
#  data above.
plot_data_2 = plot_data[plot_data$rurality == 'diff', ]
p2 = plot_data_2 %>%
  mutate(est = 100*est, lwr = 100*lwr, upr = 100*upr) %>%
  ggplot(aes(y = division, x = est)) +
  geom_point() +
  geom_errorbarh(aes(xmin = lwr, xmax = upr)) +
```

```
  geom_vline(xintercept = 0, linetype="dashed") +
  theme_bw() +
  xlab('Difference in proportion of homes with internet access')
print(p2)
```

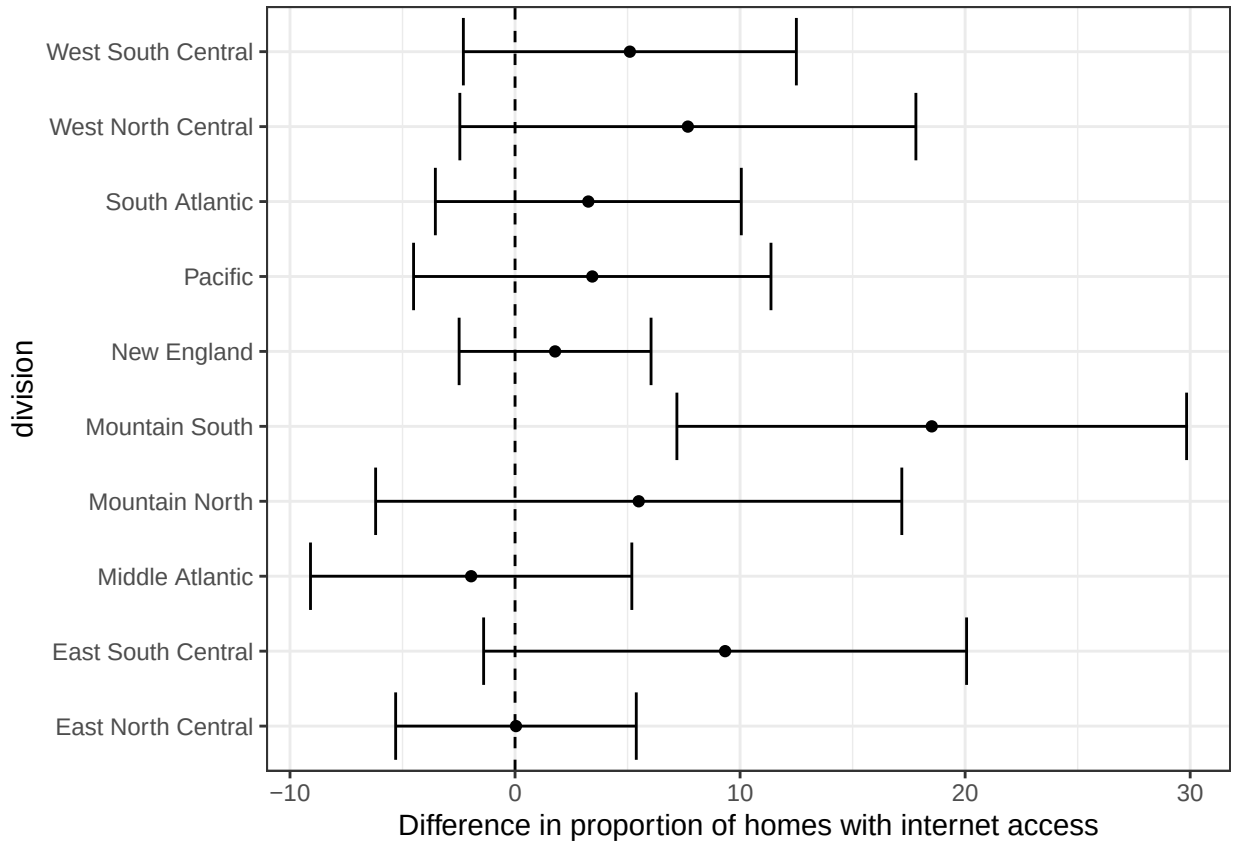

Figure 2: *Difference (urban less rural) in proportion of homes with internet access in each Census Division.*

```
# Remove comments to plot and save
ggsave('w4_p2_q2_plot.png', p2)
#p2
```