#TidyTuesday

Week 20

Moriah Taylor

5/13/2021

This Week's Data - Internet Access

The data this week comes from Microsoft by way of The Verge.

> If broadband access was a problem before 2020, the pandemic turned it into a crisis. As everyday businesses moved online, city council meetings or court proceedings became near-inaccessible to anyone whose connection couldn't support a Zoom call. Some school districts started providing Wi-Fi hotspots to students without a reliable home connection. In other districts, kids set up in McDonald's parking lots just to get a reliable enough signal to do their homework. After years of slowly widening, the broadband gap became impossible to ignore.

Note that to map the areas you will likely need to geocode by zipcode, which can be accomplished with the `{zipcodeR}` package. The `{tigris}` can also help with Census shapefiles.

`zipcodeR`

> `zipcodeR` is an R package that makes working with ZIP codes in R easier. It provides data on all U.S. ZIP codes using multiple open data sources, making it easier for social science researchers and data scientists to work with ZIP code-level data in data science projects using R.

> The latest update to zipcodeR includes new functions for searching ZIP codes at various geographic levels & geocoding.

Load Packages

```r
#load packages
library(tidytuesdayR) #tidy tuesday data
library(tidyverse) #wrangling
library(lubridate) #wrangling
library(extrafont) #fonts
library(showtext) #fonts
library(ragg)  #save as png
library(rmarkdown) #markdown
library(maps)  #geocoding
library(ggmap)  #geocoding
library(ggthemes) #custom plot theme
library(viridis)  #colorblind-friendly palette
library(sf)  #geocoding
library(tidycensus)  #geocoding
```

Load the Data

```
tuesdata <- tidytuesdayR::tt_load(2021, week = 20)

##
##  Downloading file 1 of 2: `broadband.csv`
##  Downloading file 2 of 2: `broadband_zip.csv`
```

Access the Data

```
speeds <- tuesdata$broadband
zip_codes <- tuesdata$broadband_zip
```

Wrangling

```
#rename columns
names(speeds) <- c("state", "county_id", "subregion", "broadband_fcc", "broadband_usage")
zip_codes <- zip_codes %>% rename(state = ST,
                                  subregion = "COUNTY NAME",
                                  county_id = "COUNTY ID",
                                  zipcode = "POSTAL CODE",
                                  broadband_usage = "BROADBAND USAGE")

#tidy county names into common format
speeds$subregion <- word(speeds$subregion, 1)
speeds$subregion <- tolower(speeds$subregion)
zip_codes$subregion <- word(zip_codes$subregion, 1)
zip_codes$subregion <- tolower(zip_codes$subregion)

#changing usage column to numeric
speeds$broadband_usage <- as.numeric(speeds$broadband_usage)
#creating percentage column
speeds <- speeds %>%
  mutate(broadband_perc = broadband_usage * 100)

#subset data for states of interest
nc <- speeds %>%
  filter(state == "NC")

ohio <- speeds %>%
  filter(state == "OH")

#special case: van wert county
#need to change for merge
ohio[81,3] <- "van wert"

#getting map data for each state that is ggplot-friendly
nc_county <- map_data('county', 'north carolina')
ohio_county <- map_data('county', 'ohio')

nc_state <- map_data('state', 'north carolina')
ohio_state <- map_data('state', 'ohio')

#joining data on common counties
```

```
nc_data <- merge(nc_county, nc, on="subregion")
ohio_data <- merge(ohio_county, ohio, on="subregion")


nc_data <- nc_data[order(-nc_data[,5]),]
ohio_data <- ohio_data[order(-ohio_data[,5]),]
```

Plot Stylizing

```
font_add(family = "regular", "Merriweather-Regular.ttf")
showtext_auto()


#themes
my_theme <- theme(
  # title, subtitle, caption
  plot.title = element_text(family="regular", size=55, color="white", vjust=0.5, hjust=0.5),
  plot.subtitle = element_blank(),
  plot.caption = element_text(family="regular", size=25, color="white", hjust=1),

  # panel and plot background
  panel.grid.major = element_blank(),
  panel.grid.minor = element_blank(),
  panel.background = element_rect(fill = "black"),
  plot.background = element_rect(fill = "black"),

  # axis
  axis.title = element_blank(),
  axis.text = element_blank(),
  axis.ticks = element_blank(),

  #no legend
  legend.position = "top",
  legend.title = element_text(family="regular", size=25, color="white", vjust=0.8),
  legend.background = element_rect(fill="black"),
  legend.text = element_text(family="regular", size=18, color="white", vjust=0.8)

)
```

Making the Plots

```
nc_plot <- ggplot(nc_data, #county outline and filled
                  aes(x=long, y = lat, fill=broadband_perc, group=as.factor(subregion), color="white"))

  #state outline
  geom_polygon(color="white") +


  labs(title = "Broadband Access in North Carolina",
       fill = "% of Usage at Broadband Speed",
        caption = "Moriah Taylor | Source: #TidyTuesday | Twitter: moriah_taylor58 | GitHub: moriahtaylo
```

```r
    scale_fill_viridis() + my_theme

# save image
ggsave("nc_broadband.png",
       plot = nc_plot,
       device = agg_png(width = 10, height = 5, units = "in", res = 300))

nc_plot
```

```r
oh_plot <- ggplot(ohio_data, #county outline and filled
                  aes(x=long, y = lat, fill=broadband_perc, group=as.factor(subregion), color="white"))

  #state outline
  geom_polygon(color="white") +


  labs(title = "Broadband Access in Ohio",
       fill = "% of Usage at Broadband Speed",
        caption = "Moriah Taylor | Source: #TidyTuesday | Twitter: moriah_taylor58 | GitHub: moriahtaylo

  scale_fill_viridis() + my_theme

# save image
ggsave("oh_broadband.png",
       plot = oh_plot,
       device = agg_png(width = 6, height = 7, units = "in", res = 300))

oh_plot
```

Final Plots

Broadband Access in Ohio

% of Usage at Broadband Speed

20   40   60

Broadband Access in North Carolina

% of Usage at Broadband Speed

20   40   60

5