

# R Notebook

This is an R Markdown Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Ctrl+Shift+Enter*.

```
library(shiny)
library(haven)
library(tibble)
library(ggplot2)
library(dplyr)

##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
## 
##     filter, lag
## The following objects are masked from 'package:base':
## 
##     intersect, setdiff, setequal, union
library(tidyr)

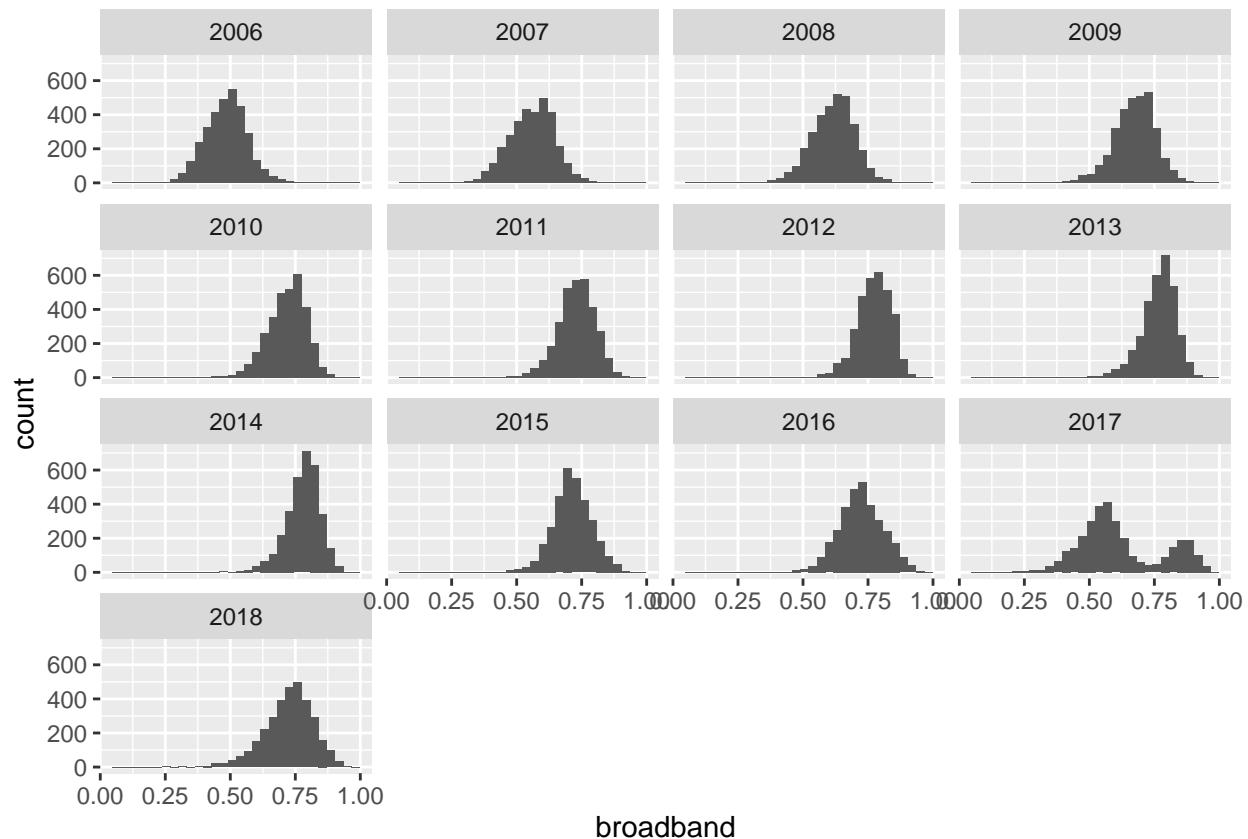
Broadband <- read_dta("broadband_county_year.dta")
broad <- as_data_frame(Broadband)

## Warning: `as_data_frame()` was deprecated in tibble 2.0.0.
## i Please use `as_tibble()` (with slightly different semantics) to convert to a
##   tibble, or `as.data.frame()` to convert to a data frame.
broadband <- as_tibble(broad)
broadband

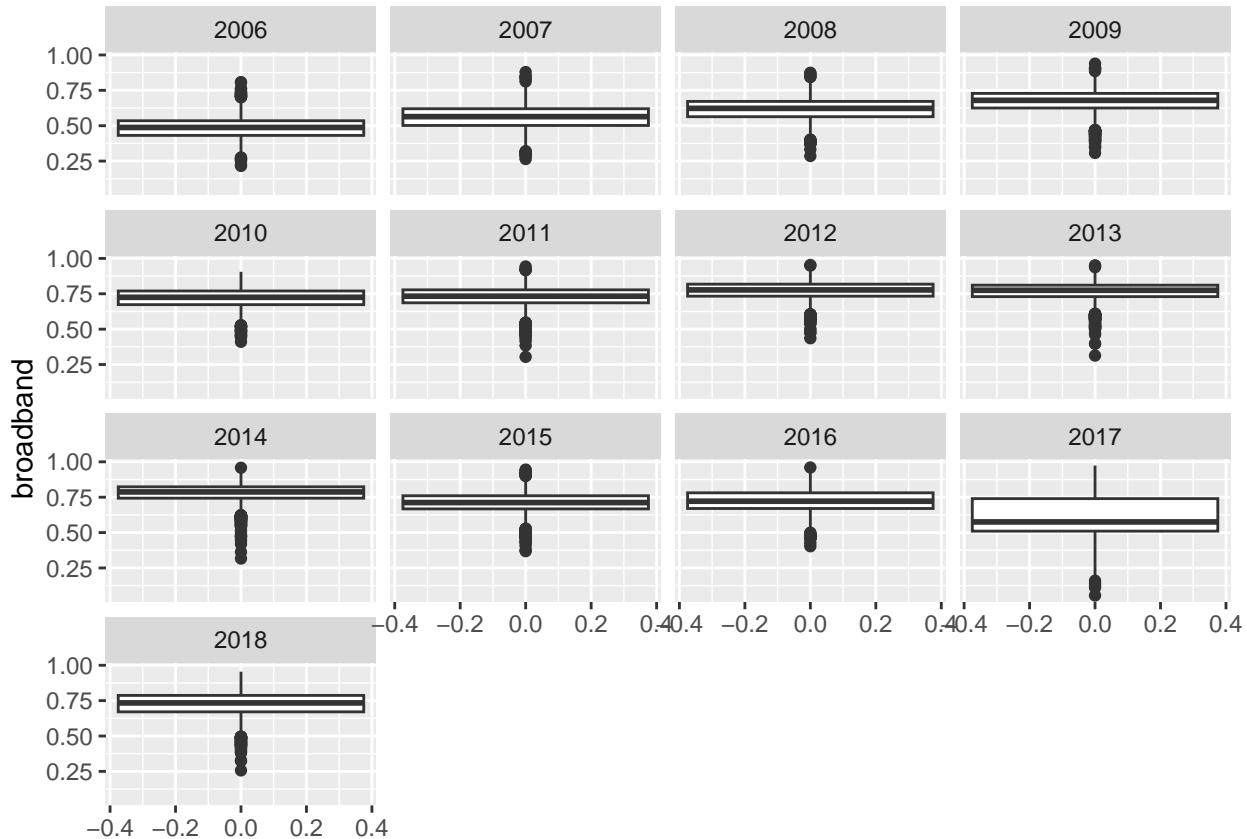
## # A tibble: 42,107 x 9
##   county  year broadband providers minority income    age population      area
##   <dbl> <dbl>     <dbl>     <dbl>     <dbl> <dbl> <dbl>     <dbl>
## 1 1001  2006     0.410     5.07     0.180  50325    17  55021.  1.70e10
## 2 1001  2007     0.452     5.09     0.205  48375    21  50226.  2.11e 9
## 3 1001  2008     0.505     8.01     0.184  50179    23  51666.  1.37e 9
## 4 1001  2009     0.597     5.77     0.242  48375    21  50085.  1.29e 9
## 5 1001  2010     0.677     6.35     0.197  50325    23  52413.  1.32e 9
## 6 1001  2011     0.723     5.97     0.234  48375    21  51305.  2.16e 9
## 7 1001  2012     0.734     7.13     0.238  67292    20  55949.  1.59e 9
## 8 1001  2013     0.729     6.80     0.224  67292    20  55882.  1.57e 9
## 9 1001  2014     0.723     5.00     0.252  67292    20  58387.  1.57e 9
## 10 1001 2015     0.724     5.40     0.217  67292    20  57264.  1.63e 9
## # ... with 42,097 more rows
```

```
# how does the distribution of broadband percentages changes over the years  
ggplot(broadband, aes(x=broadband)) + geom_histogram() +facet_wrap(~year)
```

## `stat\_bin()` using `bins = 30` . Pick better value with `binwidth` .

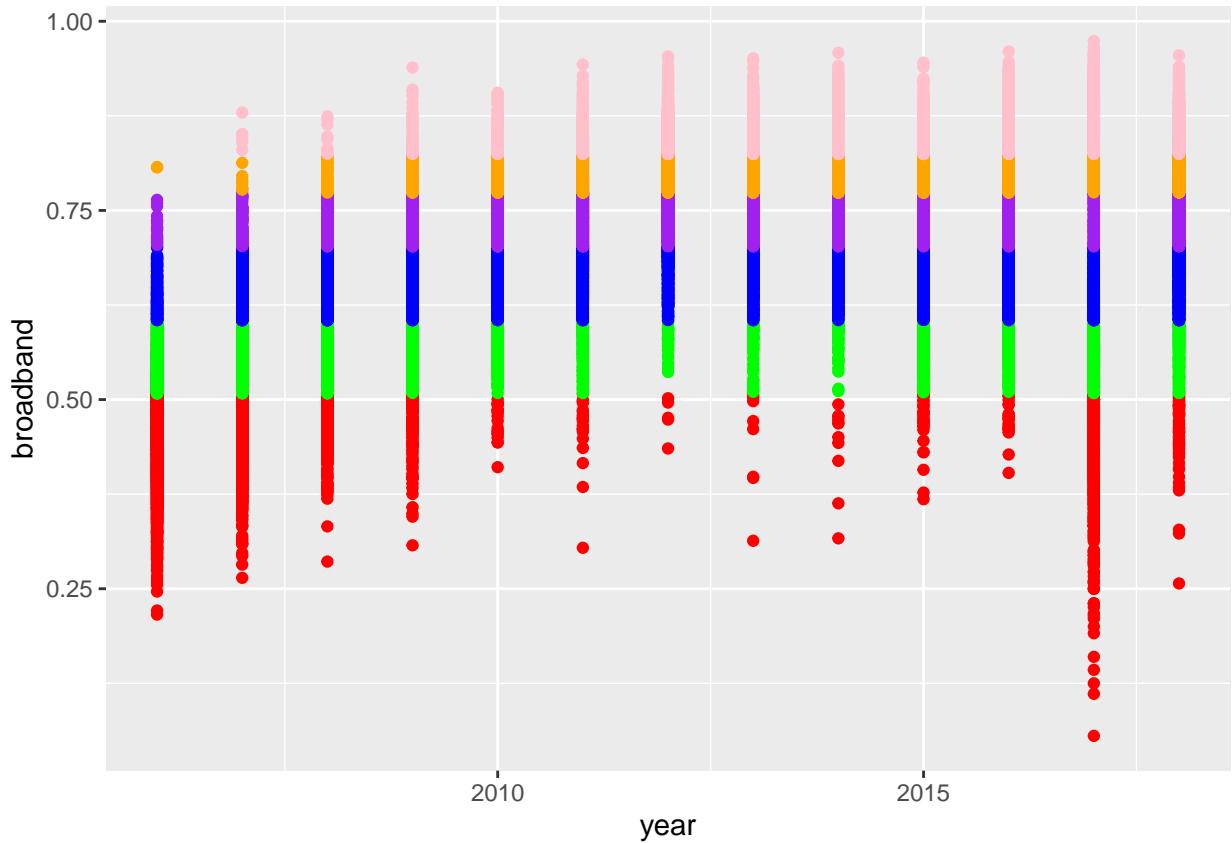


```
#how the number of counties that fall into each percentile changes over the years  
ggplot(broadband, aes(y= broadband)) + geom_boxplot() + facet_wrap(~year)
```



```
# How do the broadband percentages of different broadband percentiles (10th, 25th, 50th, 75th, 90th of
q10 <- as.numeric(quantile(broadband$broadband, 0.10))
q25 <- as.numeric(quantile(broadband$broadband, 0.25))
q50 <- as.numeric(quantile(broadband$broadband, 0.50))
q75 <- as.numeric(quantile(broadband$broadband, 0.75))
q90 <- as.numeric(quantile(broadband$broadband, 0.90))
q100 <- max(broadband$broadband)

ggplot() +
  geom_point(data = subset(broadband, broadband > 0 & broadband <= q10), aes(x = year, y = broadband),
  geom_point(data = subset(broadband, broadband > q10 & broadband <= q25), aes(x = year, y = broadband))
  geom_point(data = subset(broadband, broadband > q25 & broadband <= q50), aes(x = year, y = broadband))
  geom_point(data = subset(broadband, broadband > q50 & broadband <= q75), aes(x = year, y = broadband))
  geom_point(data = subset(broadband, broadband > q75 & broadband <= q90), aes(x = year, y = broadband))
  geom_point(data = subset(broadband, broadband > q90 & broadband <= q100), aes(x = year, y = broadband))
```

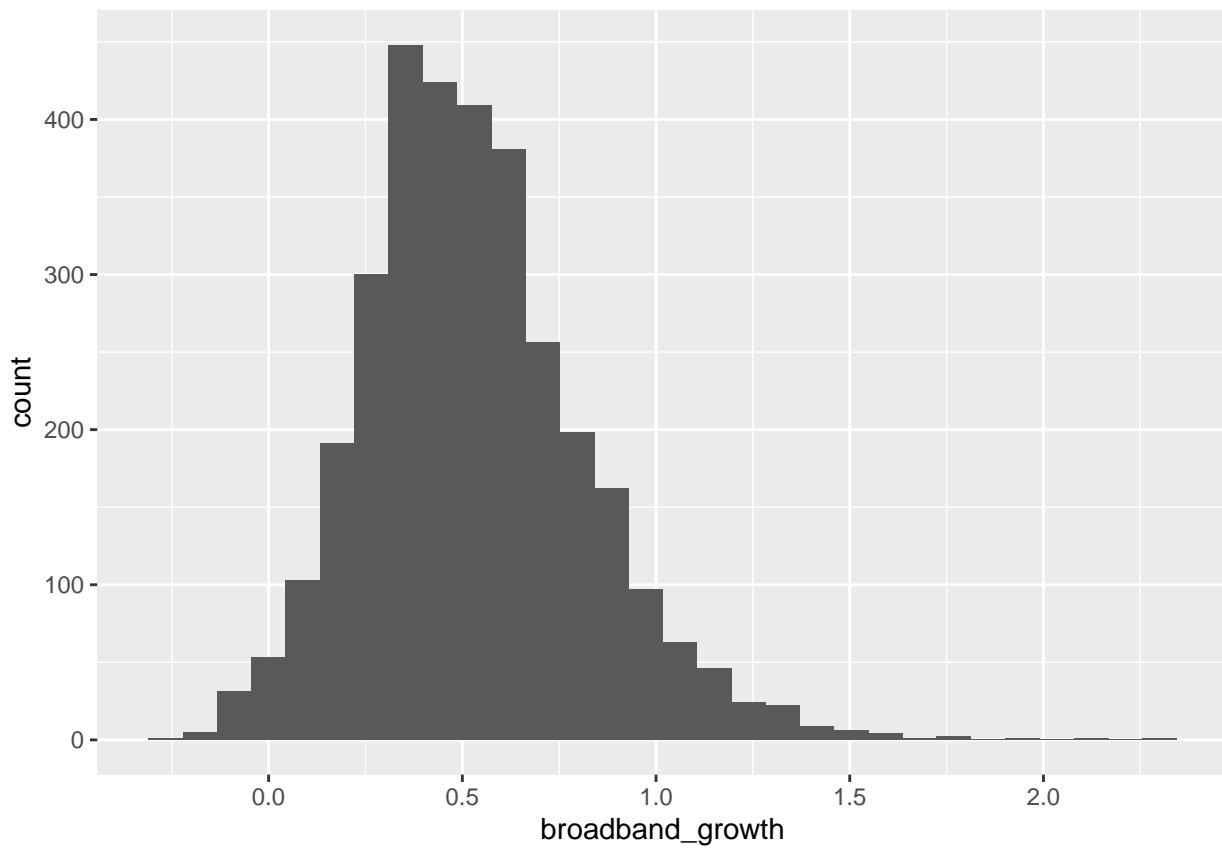


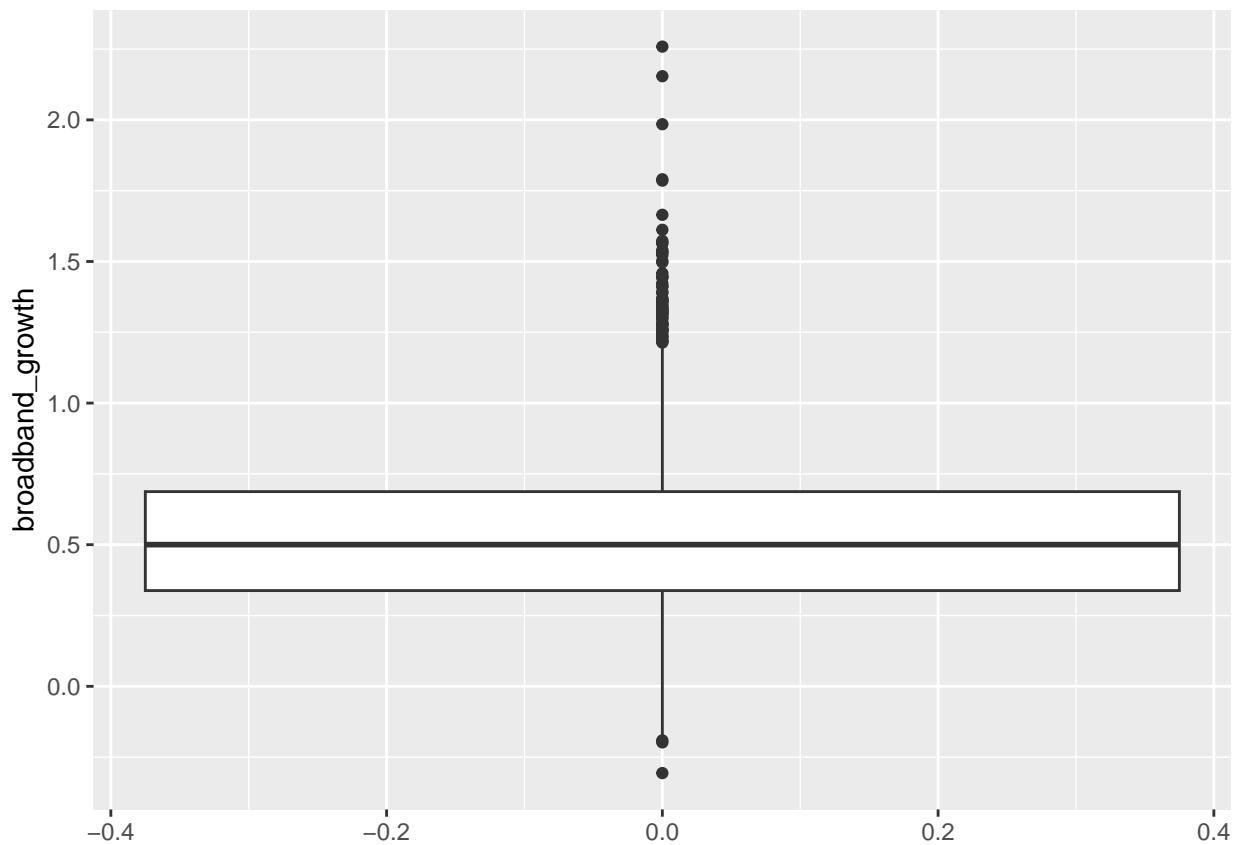
```
#Create a histogram and boxplot distribution of broadband growth rates from 2006 to 2016
filtered_data <- broadband[, c("county", "year", "broadband")]

pivot_table <- pivot_wider(filtered_data, names_from = year, values_from = broadband)

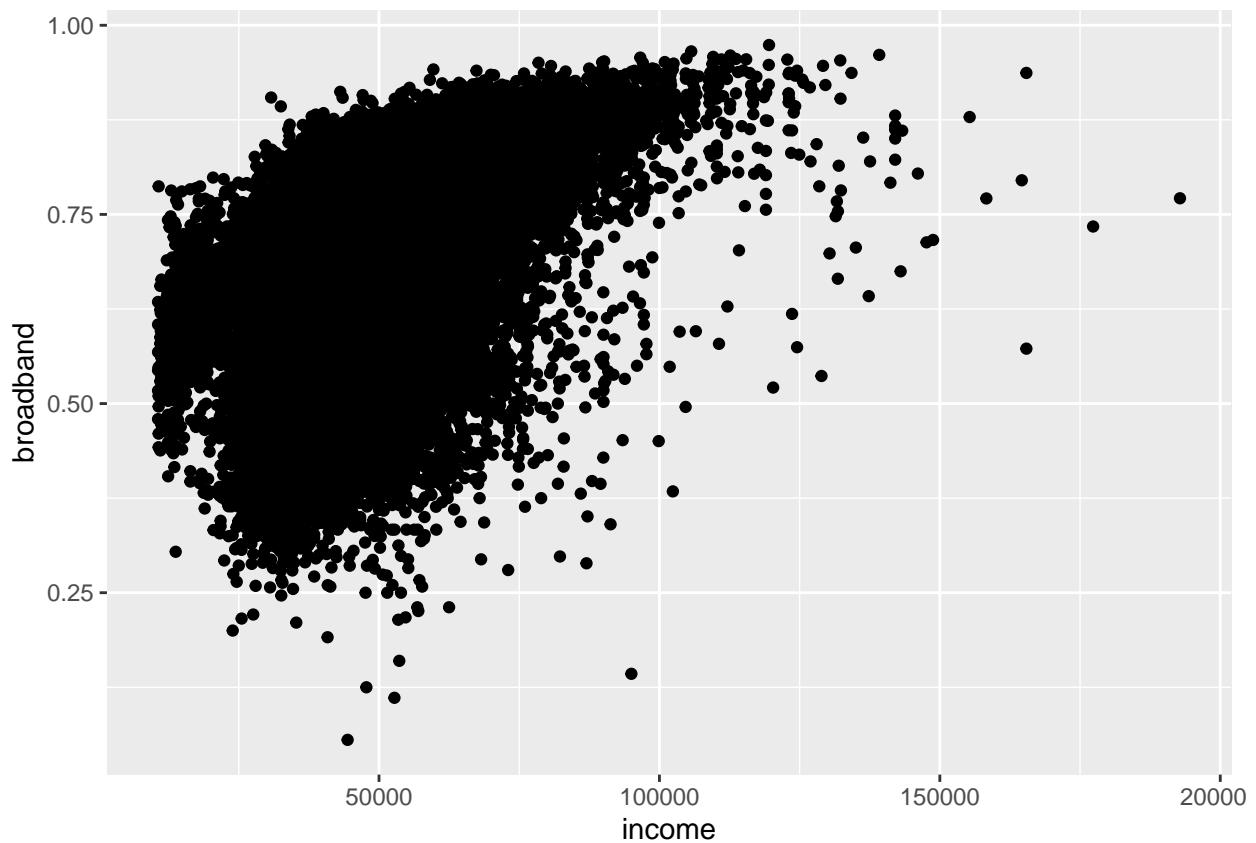
pivot_table <- as_tibble(pivot_table)
pivot_table$broadband_growth = (pivot_table[[12]] - pivot_table[[2]])/(pivot_table[[2]])
ggplot(pivot_table, aes(x=broadband_growth)) + geom_histogram()

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

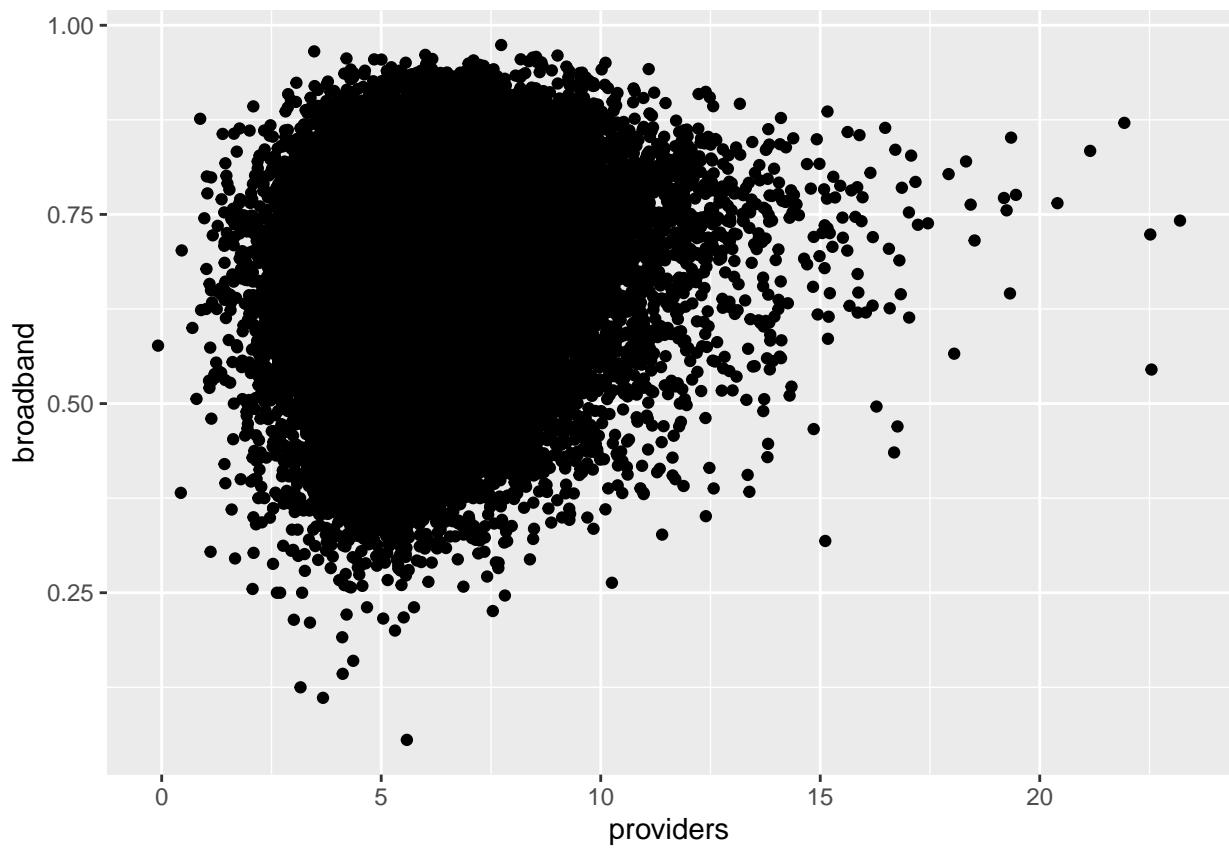




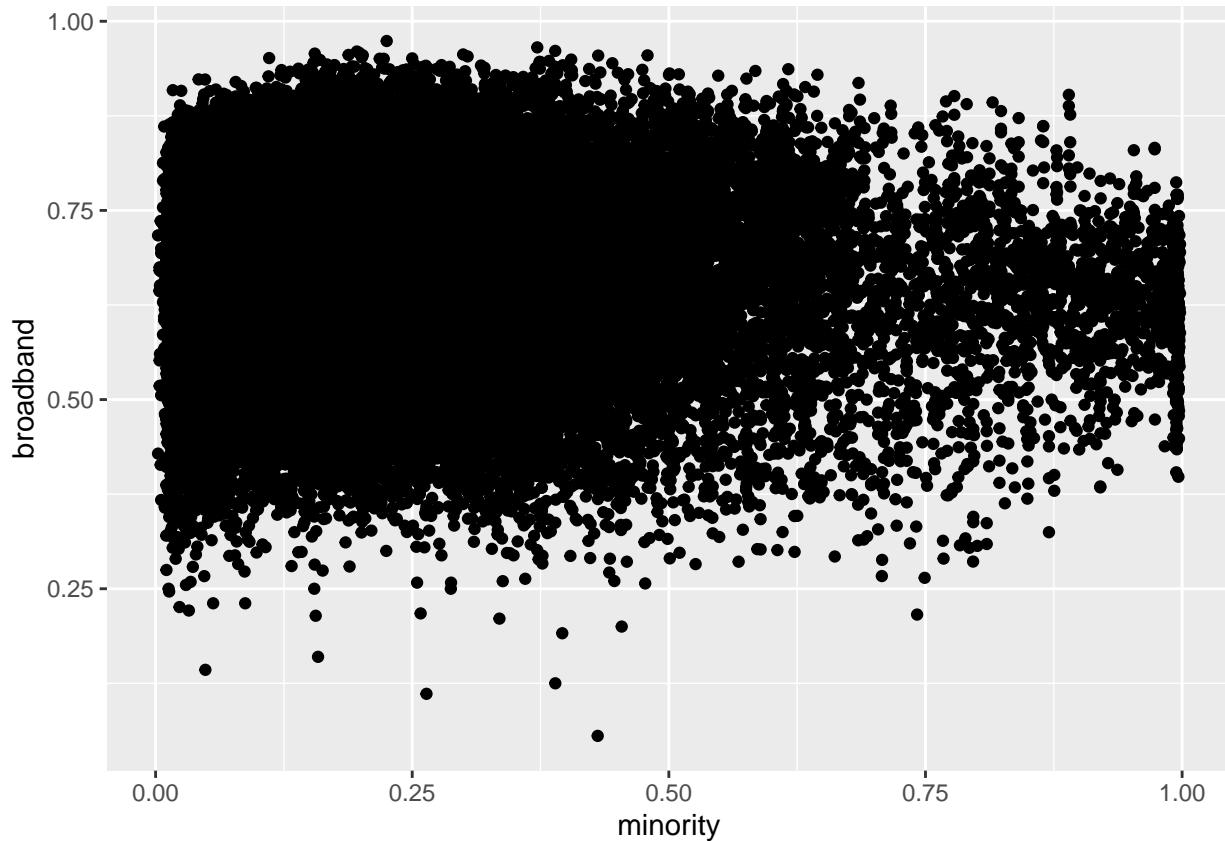
```
# how does a county's income affect the share of county with broadband (broadband percentage)?
ggplot(broadband, aes(x=income, y=broadband)) + geom_point()
```



```
# how does a county's number of providers affect the share of county with broadband(broadband percentage)
ggplot(broadband, aes(x=providers, y=broadband)) + geom_point()
```



```
# how does a county's minority population affect the share of county with broadband(broadband percentage)
ggplot(broadband, aes(x=minority, y=broadband)) + geom_point()
```



Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Ctrl+Alt+I*.

When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Ctrl+Shift+K* to preview the HTML file).

The preview shows you a rendered HTML copy of the contents of the editor. Consequently, unlike *Knit*, *Preview* does not run any R code chunks. Instead, the output of the chunk when it was last run in the editor is displayed.