Exercises-1: Intro to R - NAs

Nathan Green, Imperial College London

07/09/2019

First read in the data

```
library(dataPakistan)
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.5.3
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.5.3
```

```
##
## Attaching package: 'dplyr'
```

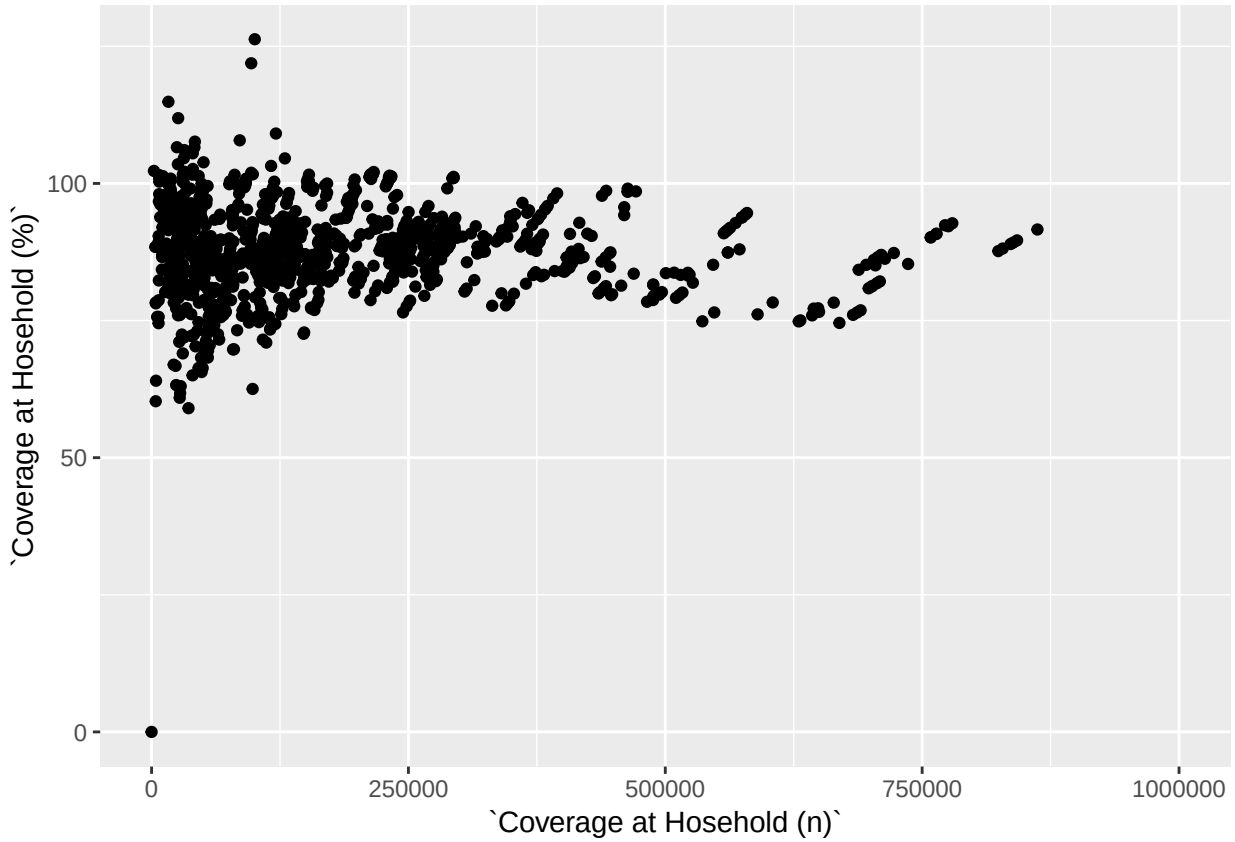
```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
file_location <- system.file("extdata", package = "dataPakistan")
dat <- readxl::read_excel(path = paste0(file_location, "/Admin-datasheet-year2018.xlsx"))
```

```
ggplot(data = dat, mapping = aes(x = `Coverage at Hosehold (n)`, y = `Coverage at Hosehold (%)`)) + geo
```

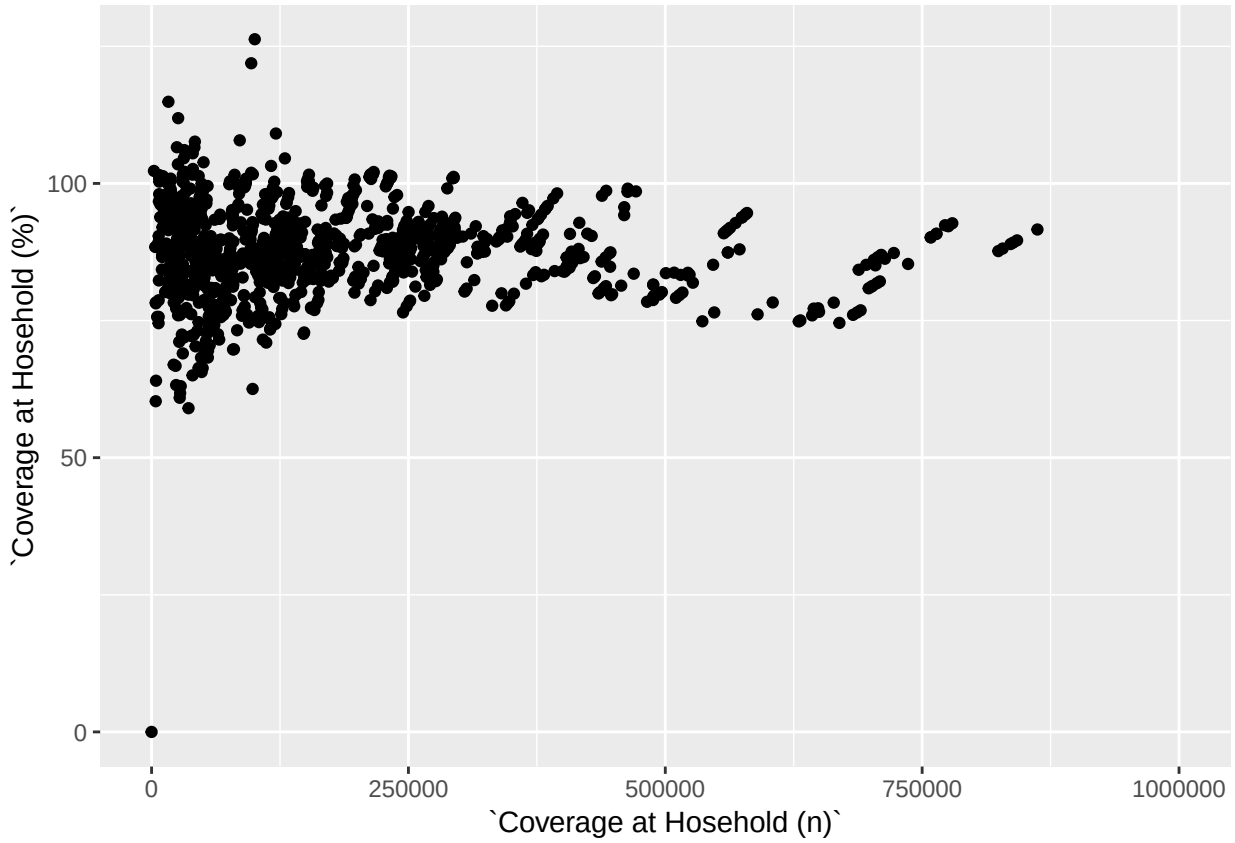
```
## Warning: Removed 21 rows containing missing values (geom_point).
```




```
my_plot <- ggplot(data = dat, mapping = aes(x = `Coverage at Hosehold (n)`, y = `Coverage at Hosehold (
```

```
my_plot + geom_point() + xlim(0, 1e+6)
```

```
## Warning: Removed 21 rows containing missing values (geom_point).
```



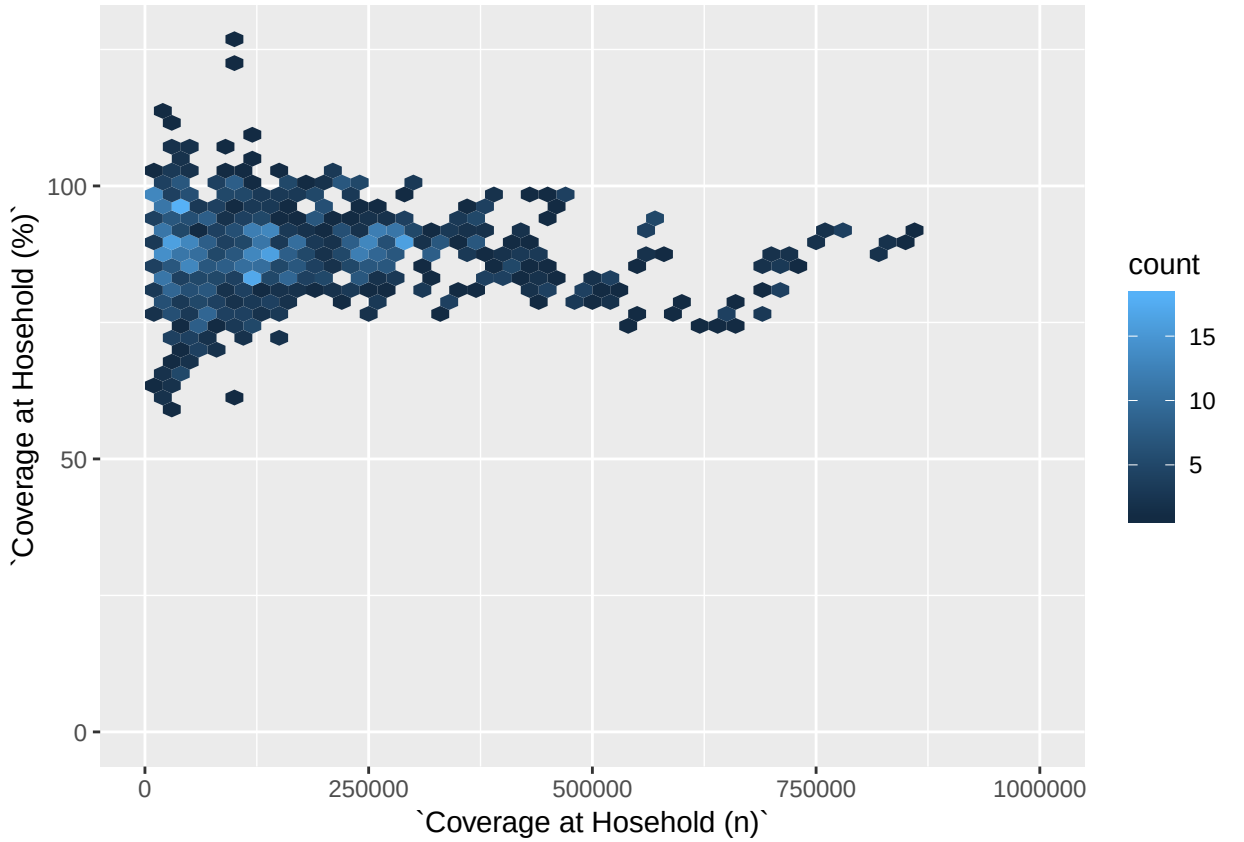

```
# install.packages("hexbin")
library(hexbin)
```

```
## Warning: package 'hexbin' was built under R version 3.5.3
```

```
my_plot + geom_hex(bins = 50) + xlim(0, 1e+6)
```

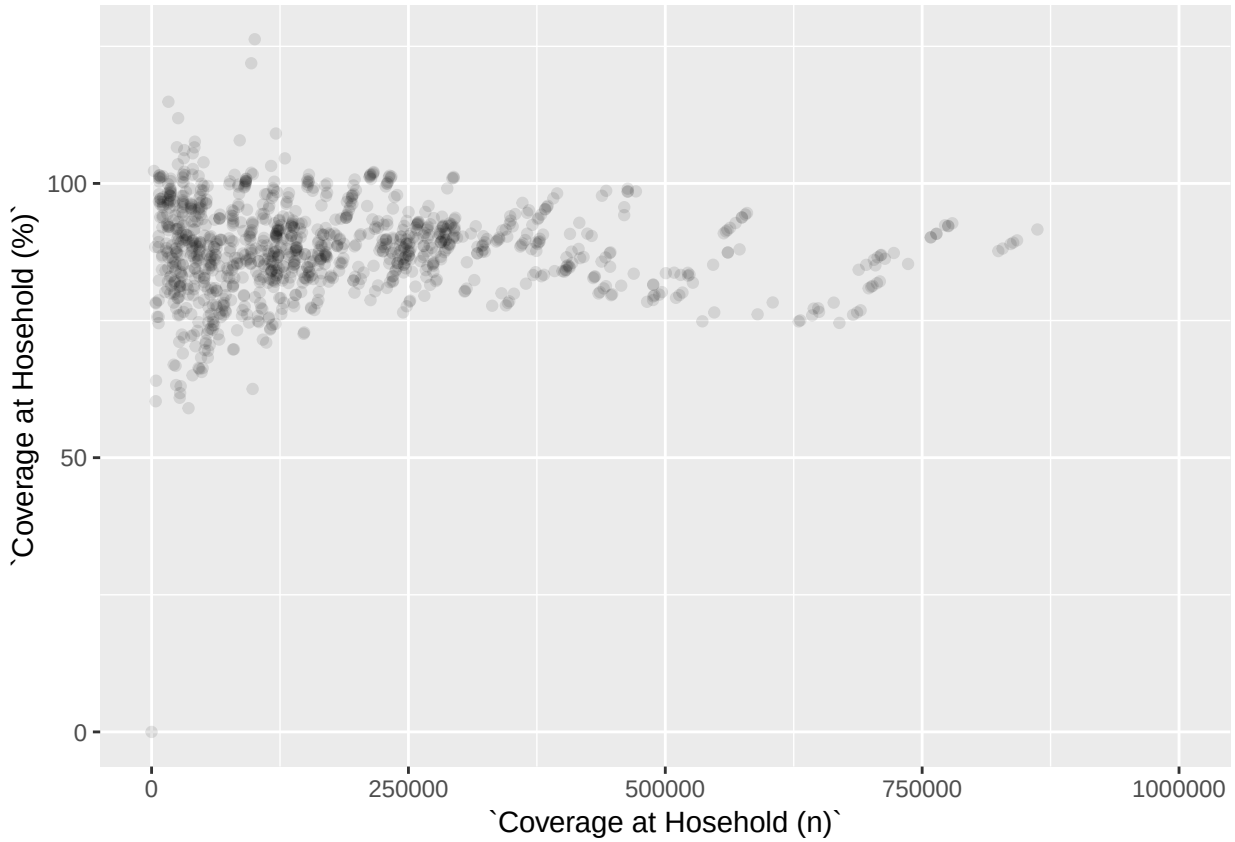
```
## Warning: Removed 21 rows containing non-finite values (stat_binhex).
```

```
## Warning: Removed 8 rows containing missing values (geom_hex).
```



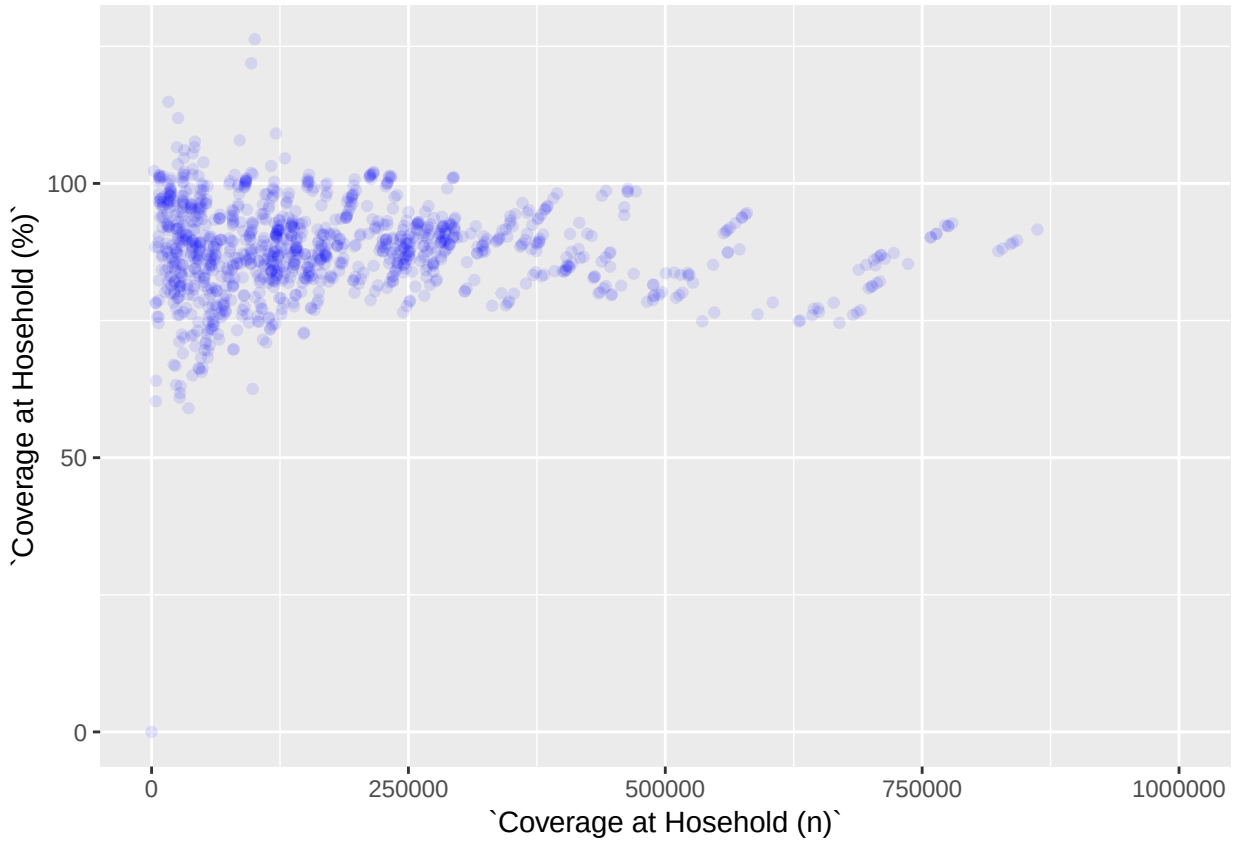

```
my_plot + geom_point(alpha = 0.1) + xlim(0, 1e+6)
```

```
## Warning: Removed 21 rows containing missing values (geom_point).
```



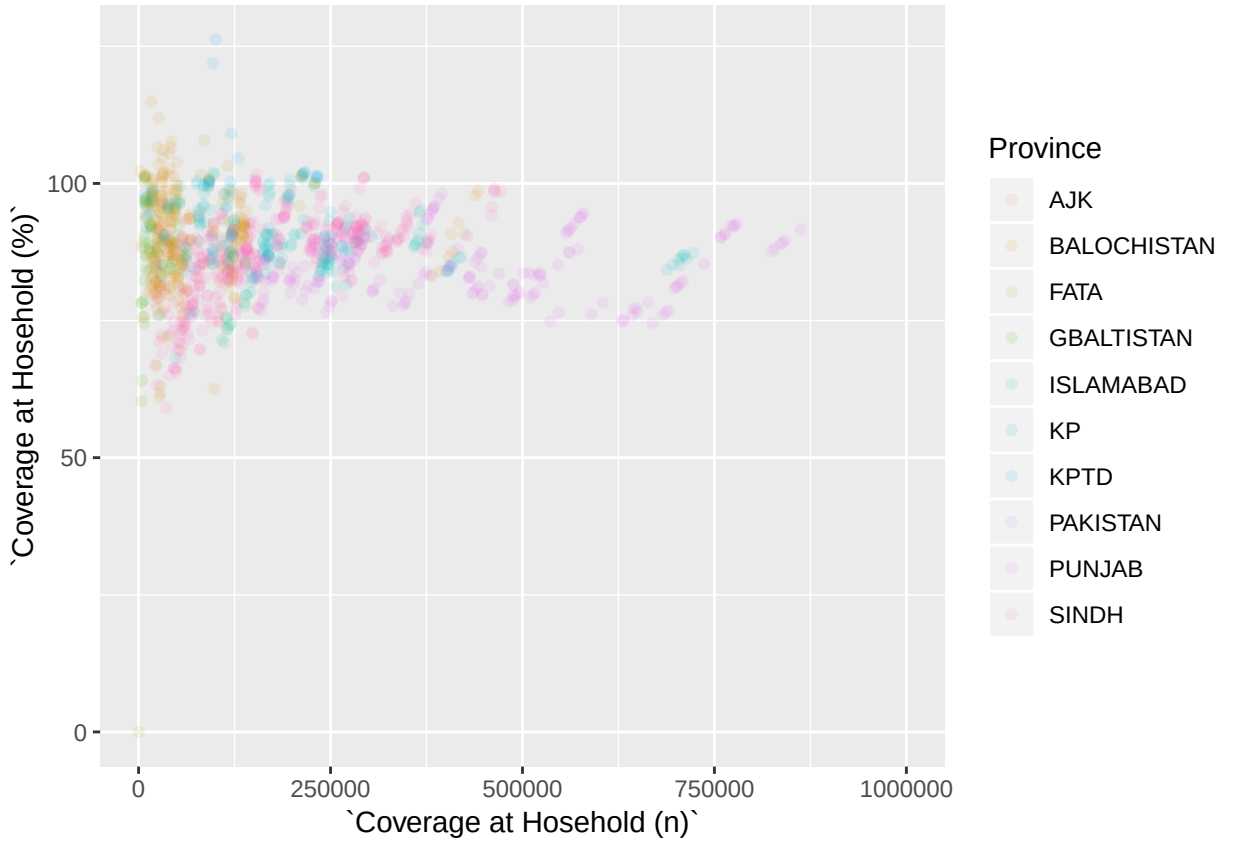

```
my_plot + geom_point(alpha = 0.1, colour = "blue") + xlim(0, 1e+6)
```

```
## Warning: Removed 21 rows containing missing values (geom_point).
```



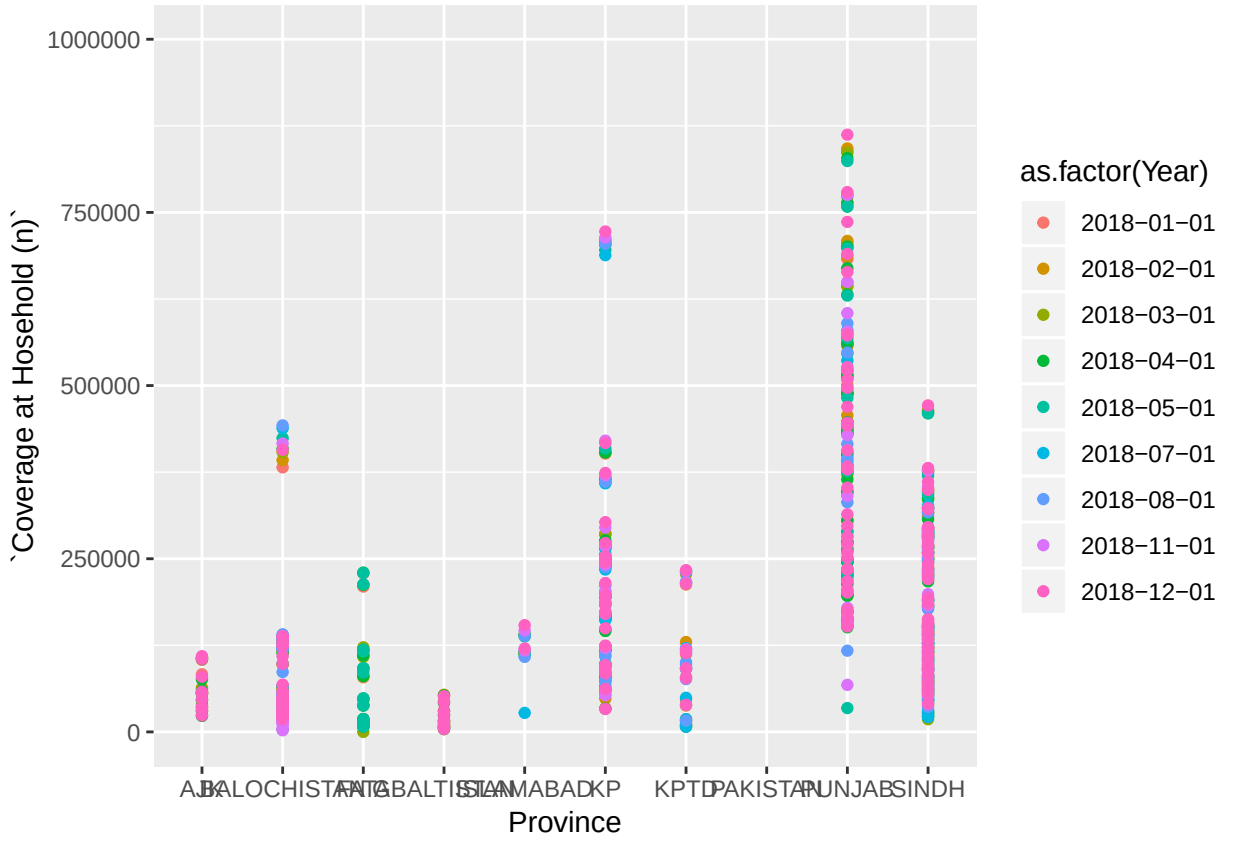

```
my_plot + geom_point(alpha = 0.1, aes(colour = Province)) + xlim(0, 1e+6)
```

```
## Warning: Removed 21 rows containing missing values (geom_point).
```



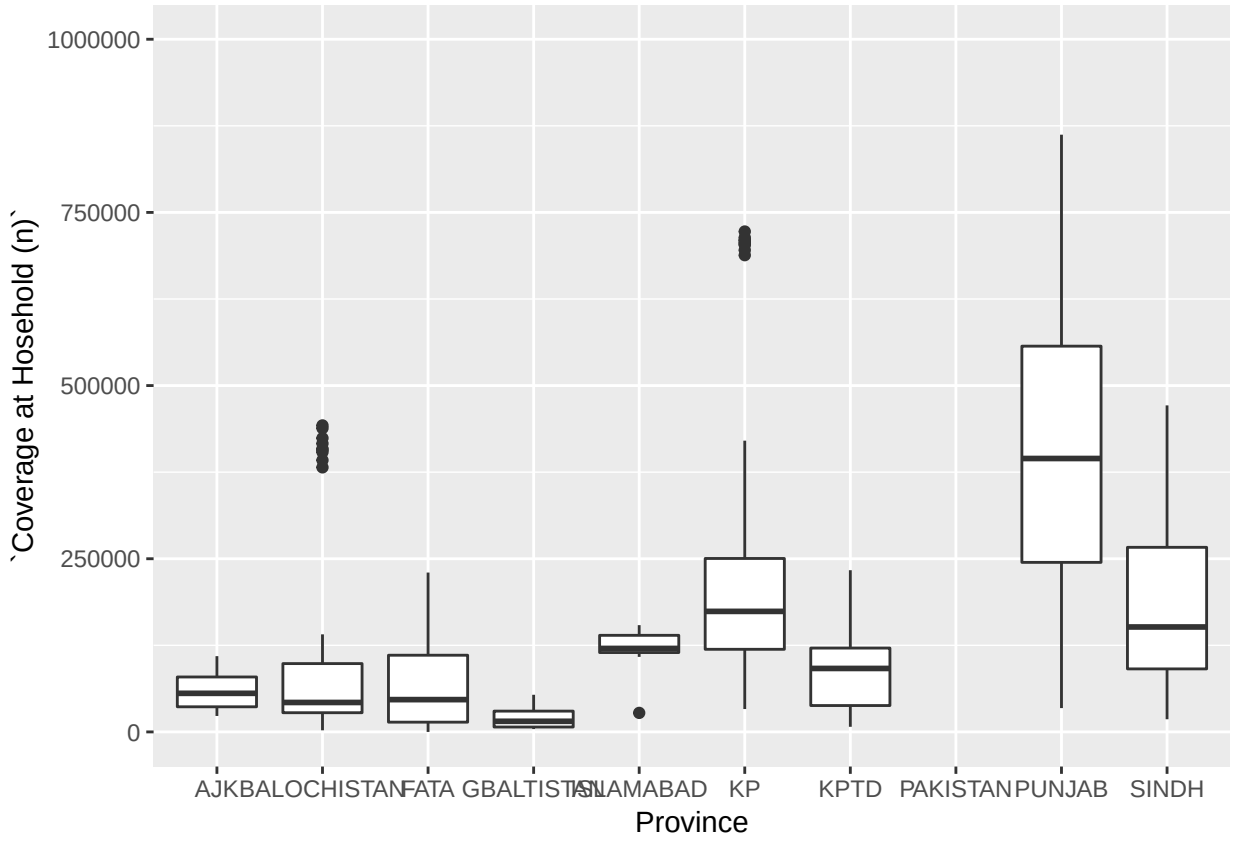

```
ggplot(data = dat,
       mapping = aes(x = Province, y = `Coverage at Hosehold (n)`)) +
  geom_point(aes(colour = as.factor(Year))) + ylim(0, 1e+6)
```

```
## Warning: Removed 21 rows containing missing values (geom_point).
```




```
ggplot(data = dat,
       mapping = aes(x = Province, y = `Coverage at Hosehold (n)`)) +
  geom_boxplot() + ylim(0, 1e+6)
```

```
## Warning: Removed 21 rows containing non-finite values (stat_boxplot).
```




```
ggplot(data = dat,
       mapping = aes(x = Province, y = `Coverage at Hosehold (n)`)) +
  geom_boxplot(alpha = 0) +
  geom_jitter(alpha = 0.3, colour = "tomato") +
  ylim(0, 1e+6)
```

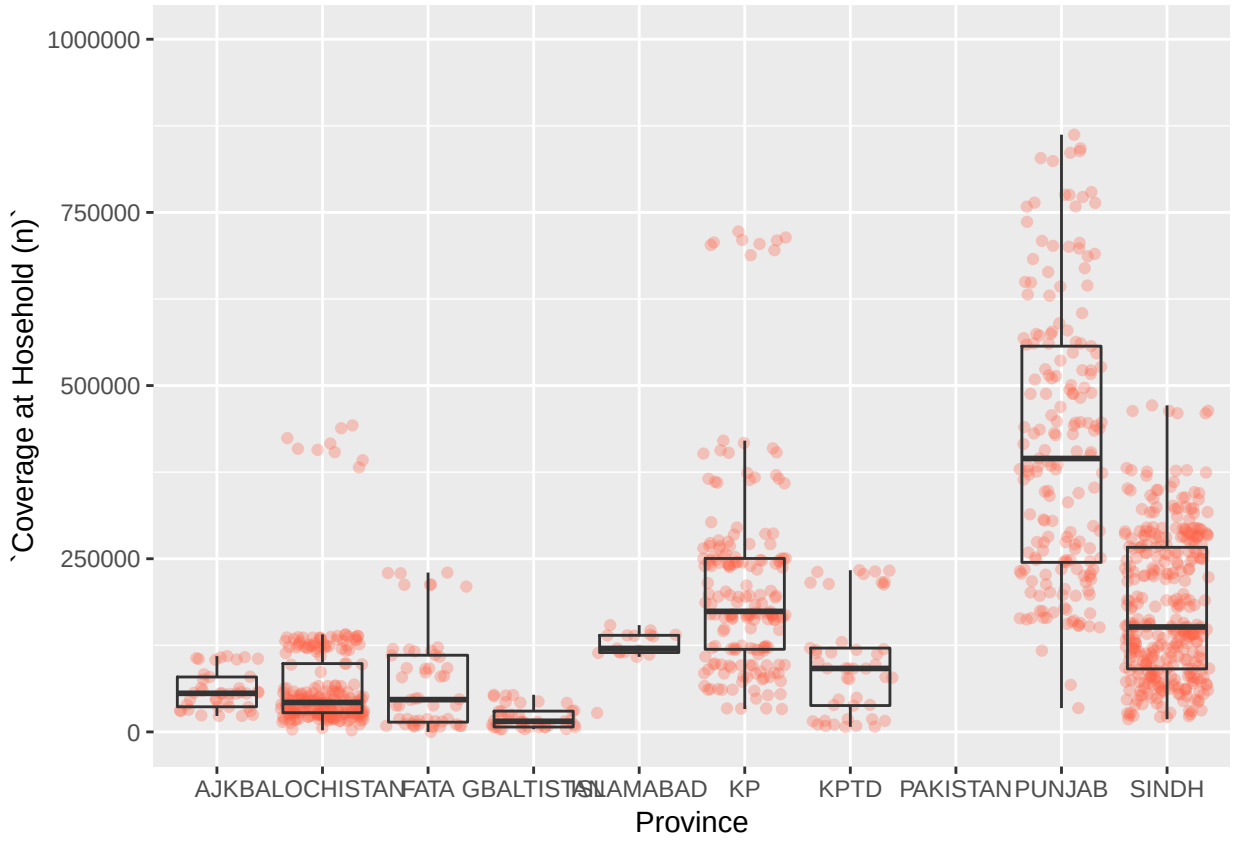
```
## Warning: Removed 21 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 21 rows containing missing values (geom_point).
```


```
ggplot(data = dat,
       mapping = aes(x = Province, y = `Coverage at Hosehold (n)`)) +
  geom_jitter(alpha = 0.3, colour = "tomato") +
  geom_boxplot(alpha = 0) +
  ylim(0, 1e+6)
```

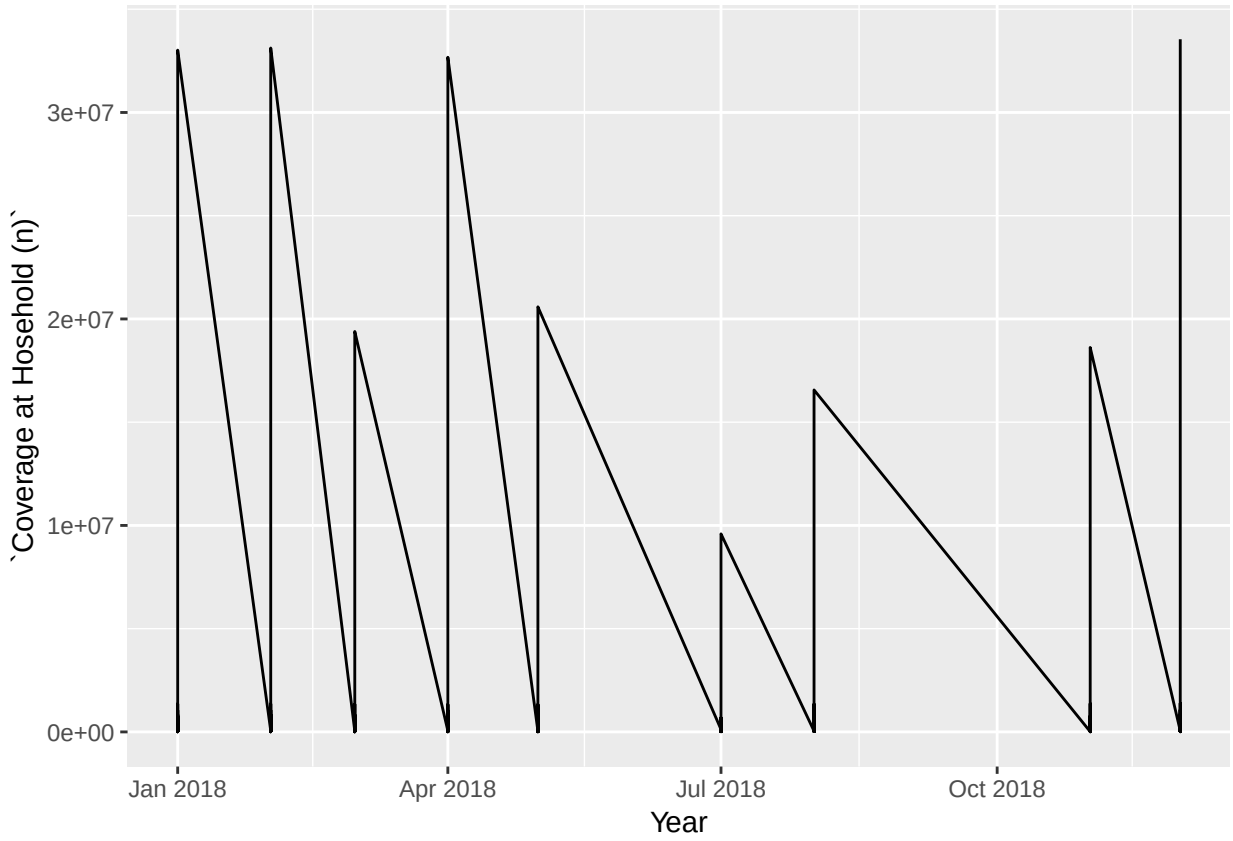
```
## Warning: Removed 21 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 21 rows containing missing values (geom_point).
```

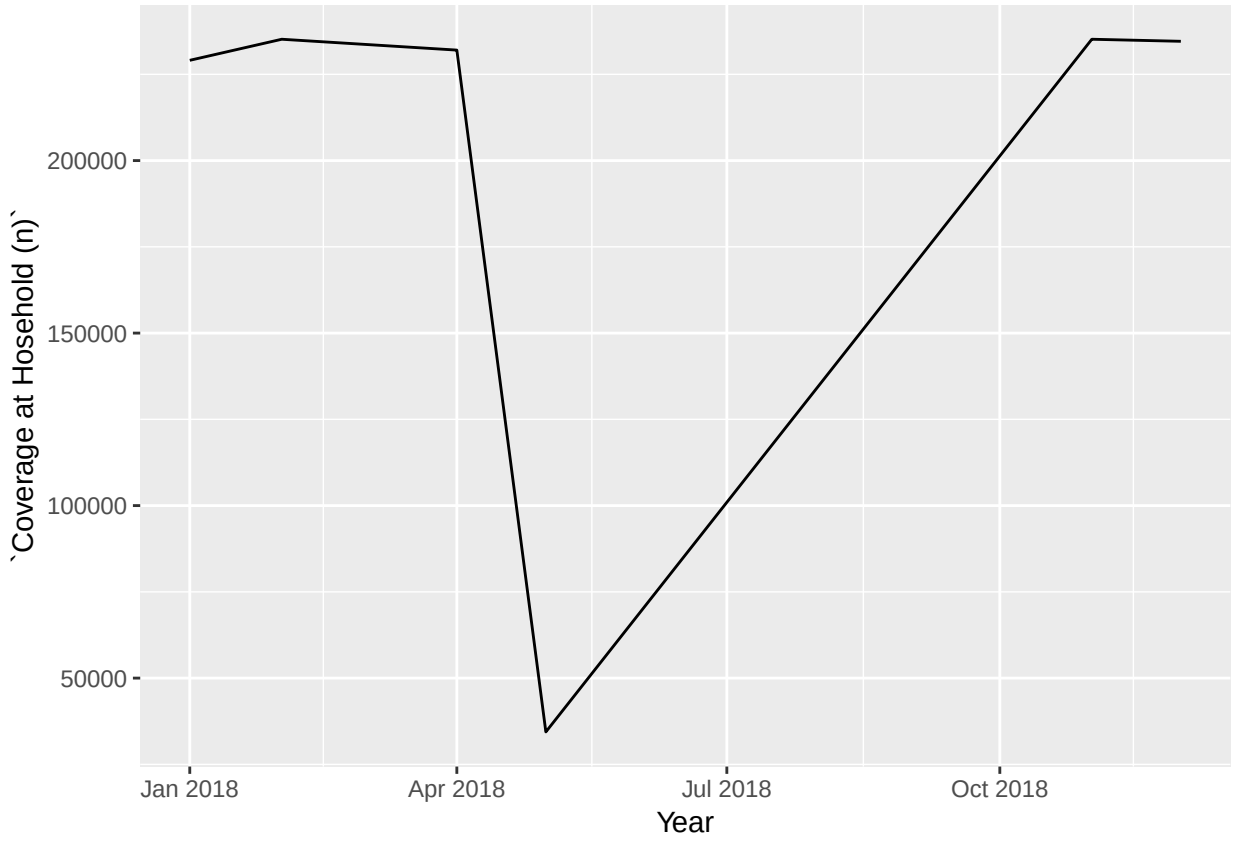


time series

```
ggplot(data = dat, aes(x = Year, y = `Coverage at Hosehold (n)`)) + geom_line()
```

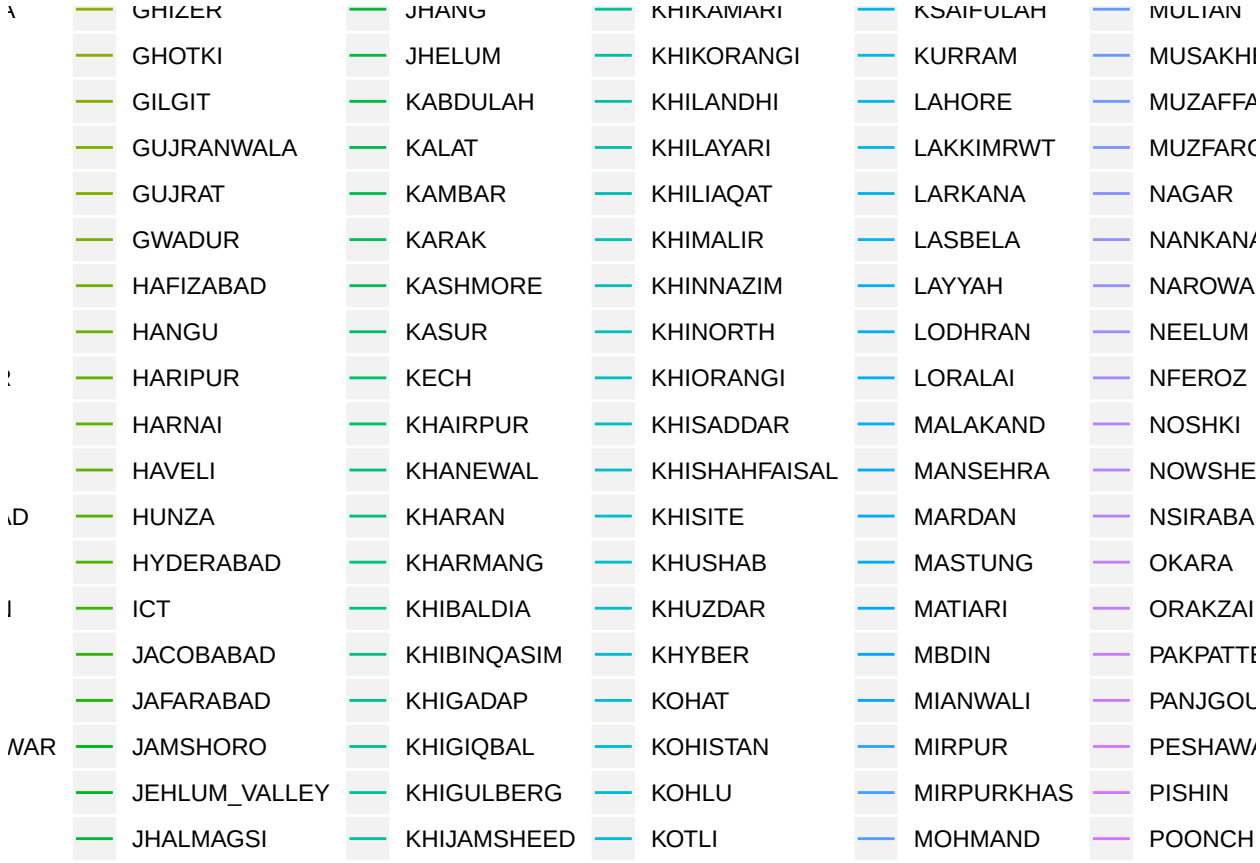



```
ggplot(data = dat[dat$District == "ATTOCK", ], aes(x = Year, y = `Coverage at Hosehold (n)`)) + geom_li
```




```
ggplot(data = dat, aes(x = Year, y = `Coverage at Hosehold (n)`, colour = District)) + geom_line() + yli
```

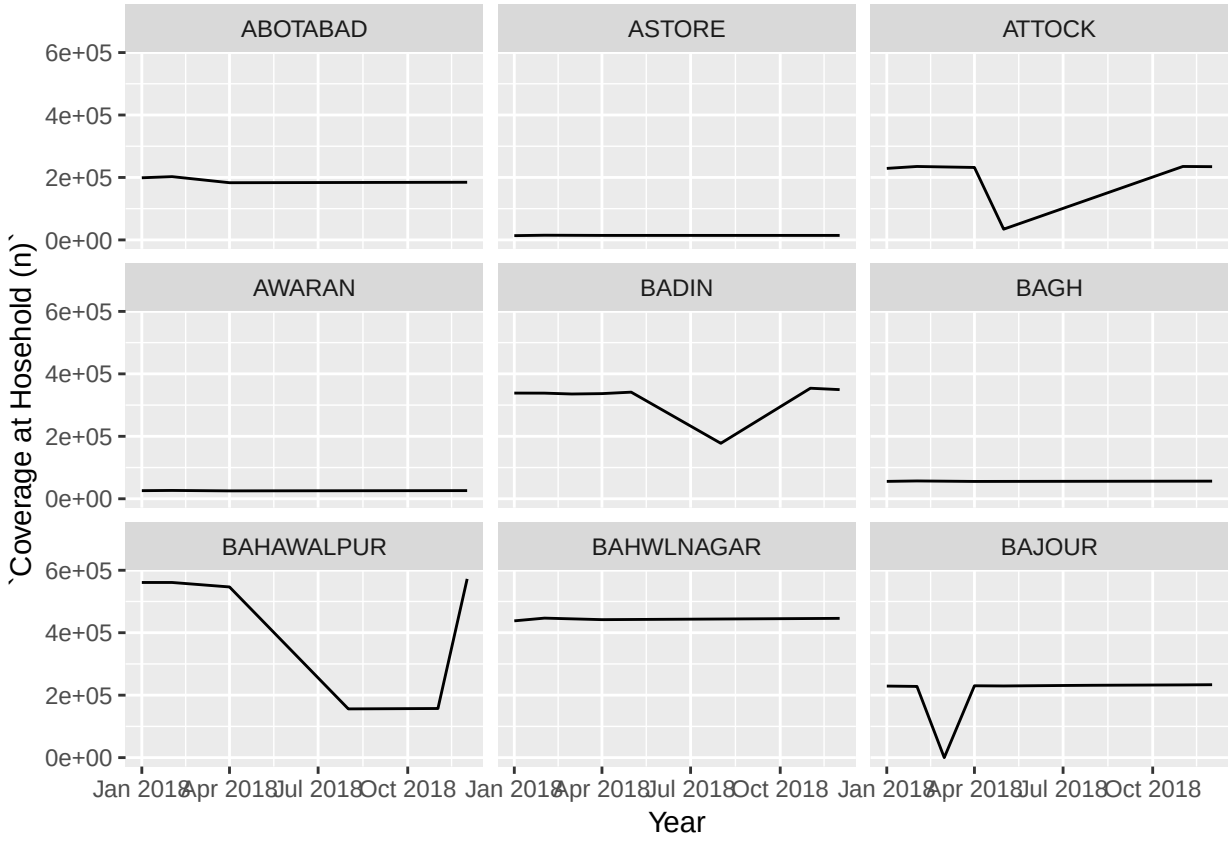
```
## Warning: Removed 21 rows containing missing values (geom_path).
```



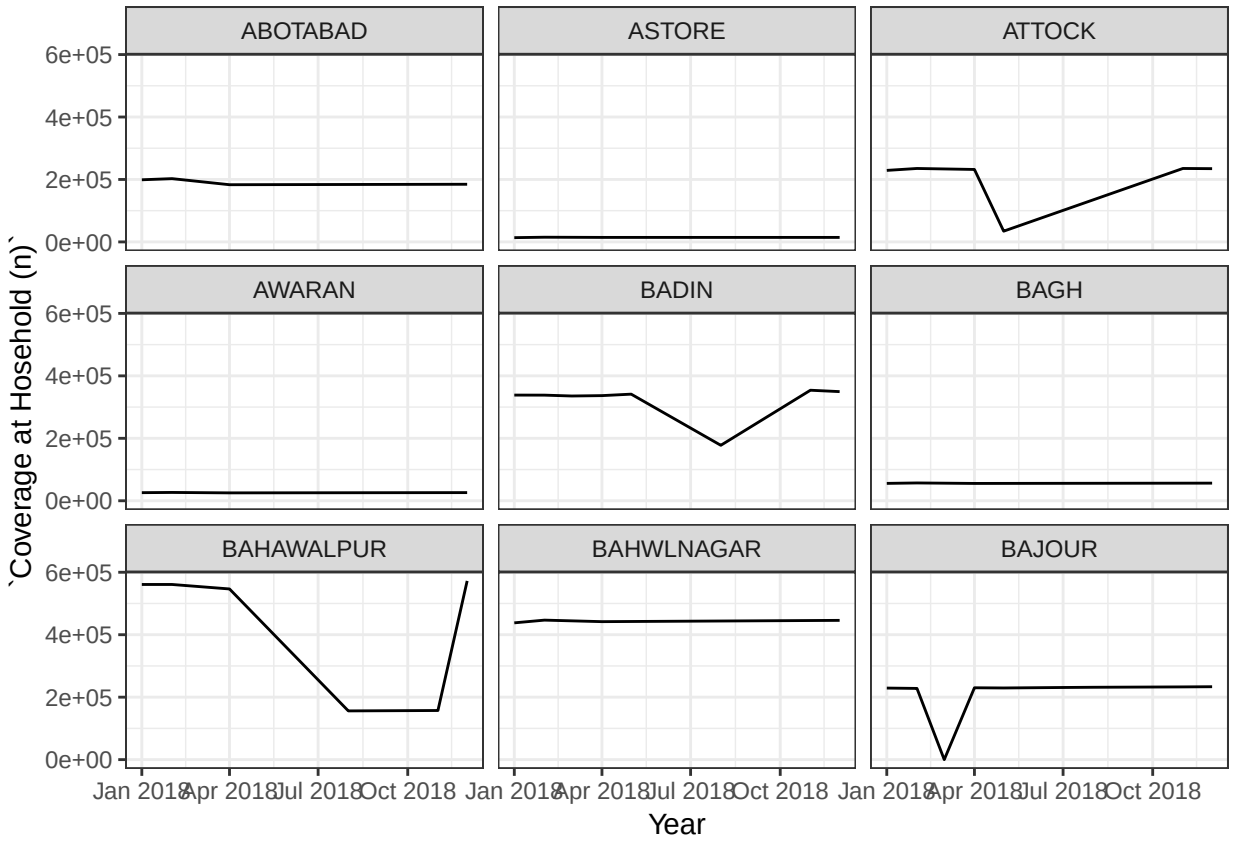
faceting

```
nms <- table(dat$District) %>% names %>% .[1:9]
ss_dat <- dat[dat$District %in% nms, ]

ggplot(data = ss_dat, mapping = aes(x = Year, y = `Coverage at Hosehold (n)`)) +
  geom_line() +
  facet_wrap(facets = vars(District))
```




```
ggplot(data = ss_dat, mapping = aes(x = Year, y = `Coverage at Hosehold (n)`)) +
  geom_line() +
  facet_wrap(facets = vars(District)) +
  theme_bw()
```



arranging and exporting plots

```
# install.packages("gridExtra")
library(gridExtra)
```

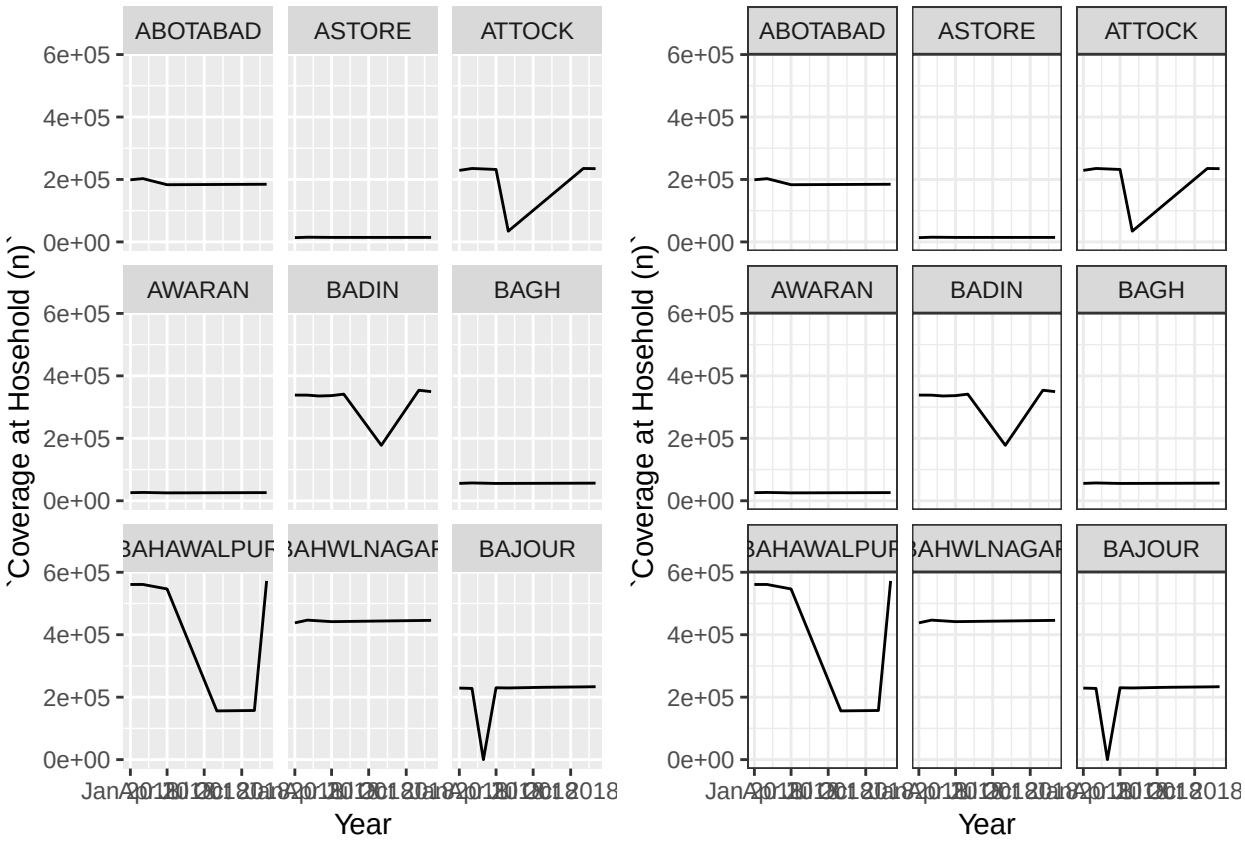
```
##
## Attaching package: 'gridExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

```
p1 <-
  ggplot(data = ss_dat, mapping = aes(x = Year, y = `Coverage at Hosehold (n)`)) +
  geom_line() +
  facet_wrap(facets = vars(District))

p2 <-
  ggplot(data = ss_dat, mapping = aes(x = Year, y = `Coverage at Hosehold (n)`)) +
  geom_line() +
  facet_wrap(facets = vars(District)) +
  theme_bw()

final_plot <- grid.arrange(p1, p2, ncol = 2)
```




```
ggsave("final_plot.png", final_plot, width = 10, dpi = 300)
```