Exercises-1: ggplot2

Nathan Green, Imperial College London

07/09/2019

First read in the data

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

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

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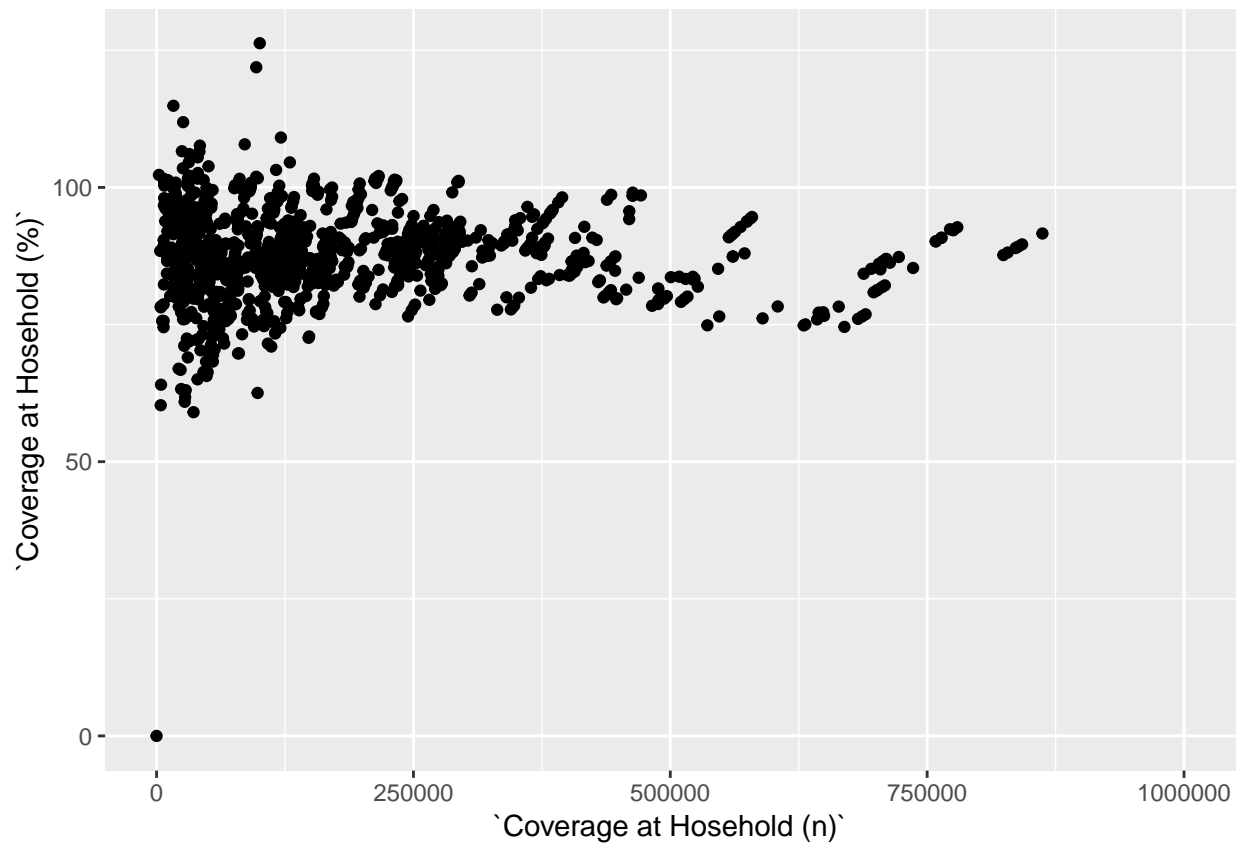
```r
file_location <- system.file("extdata", package = "dataPakistan")
dat <- readxl::read_excel(path = paste0(file_location, "/Admin-datasheet-year2018.xlsx"))
```

```r
ggplot(data = dat, mapping = aes(x = `Coverage at Hosehold (n)`, y = `Coverage at Hosehold (%)`)) + geo
```
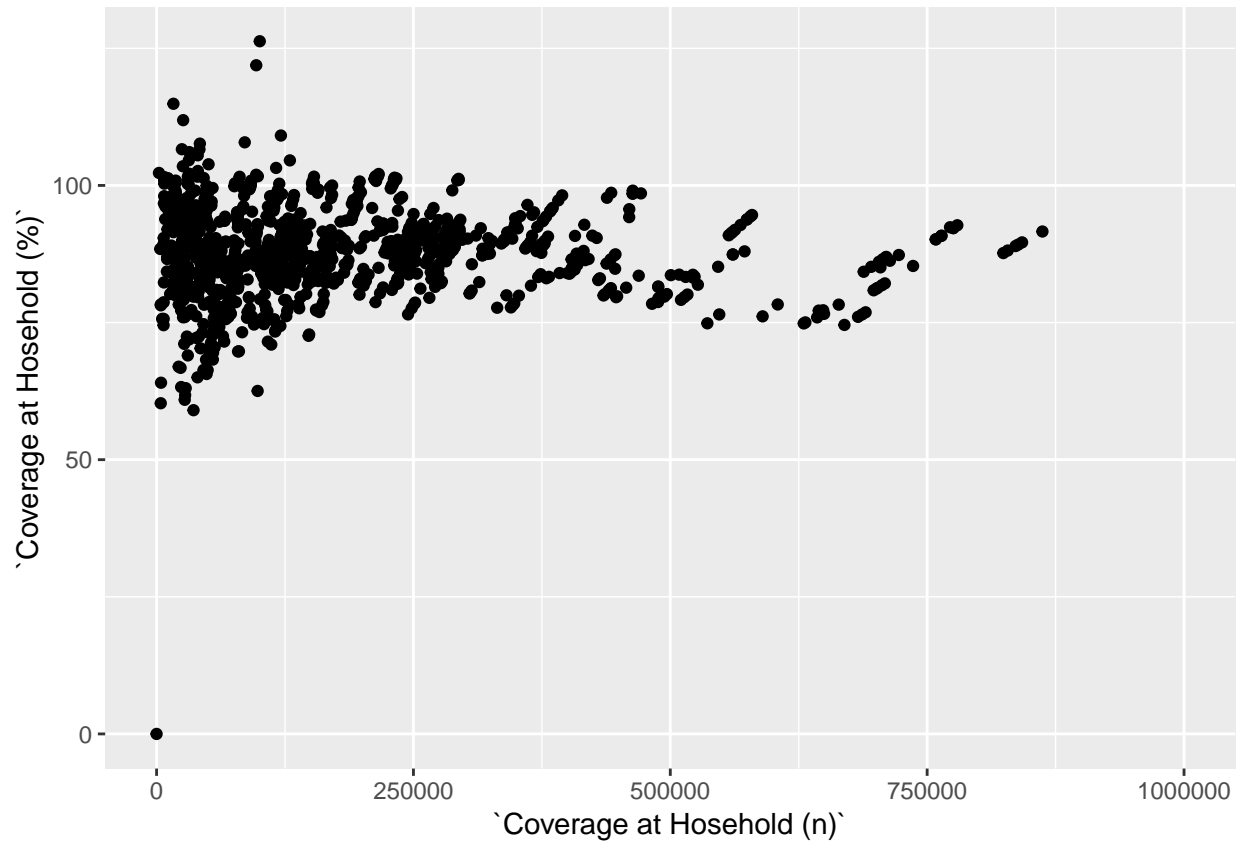
```
## Warning: Removed 21 rows containing missing values (geom_point).
```

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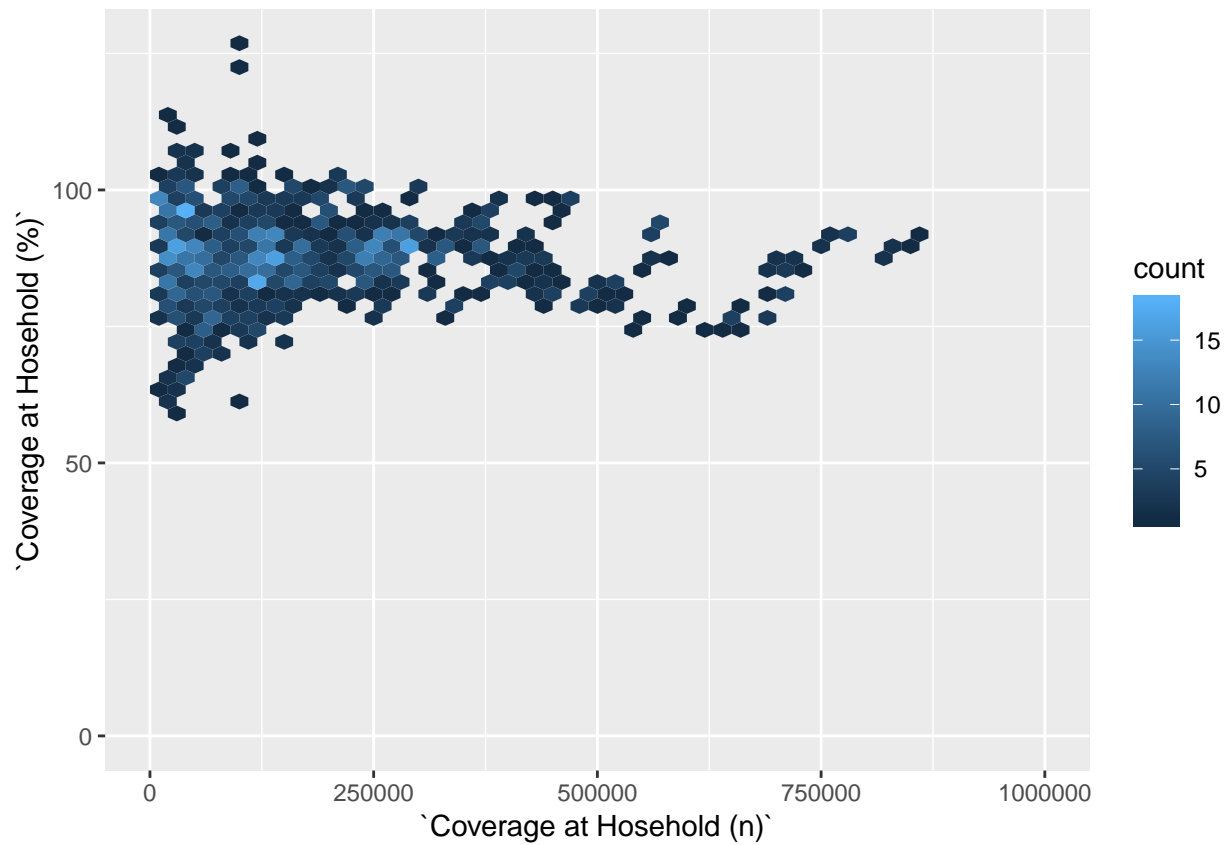
```
# install.packages("hexbin")
library(hexbin)
```

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

```
my_plot + geom_hex(bins = 50) + xlim(0, 1e+6)
```
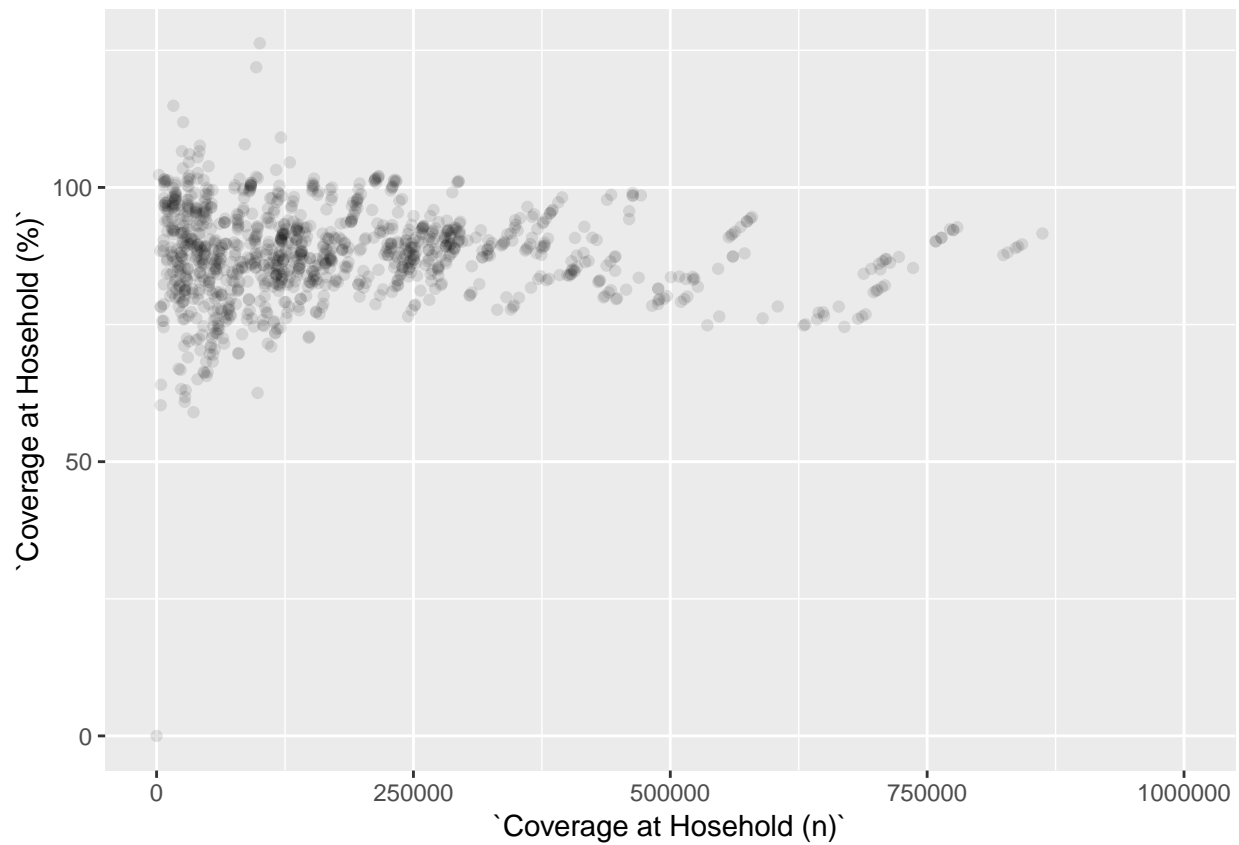
```
## Warning: Removed 21 rows containing non-finite values (stat_binhex).
```

```
## Warning: Removed 8 rows containing missing values (geom_hex).
```
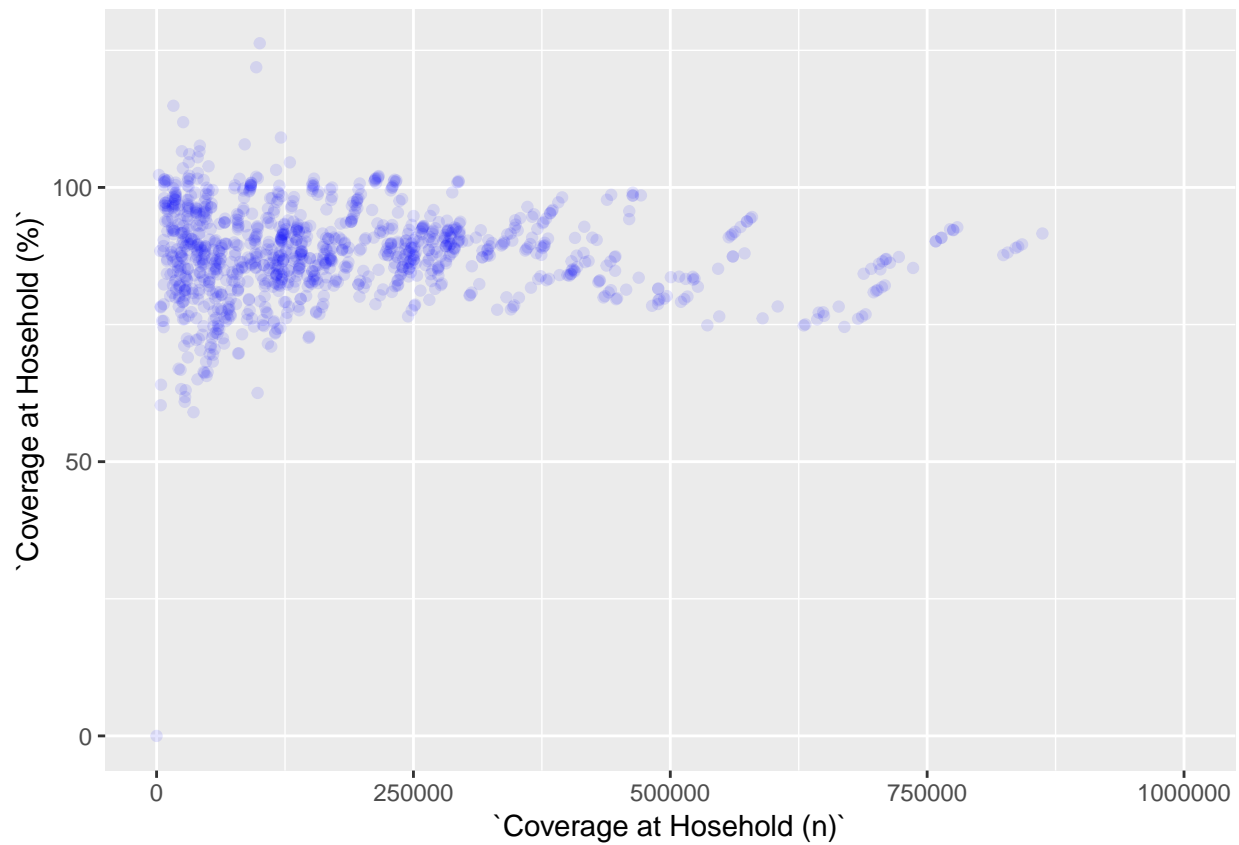
```
my_plot + geom_point(alpha = 0.1) + xlim(0, 1e+6)
```

```
## Warning: Removed 21 rows containing missing values (geom_point).
```
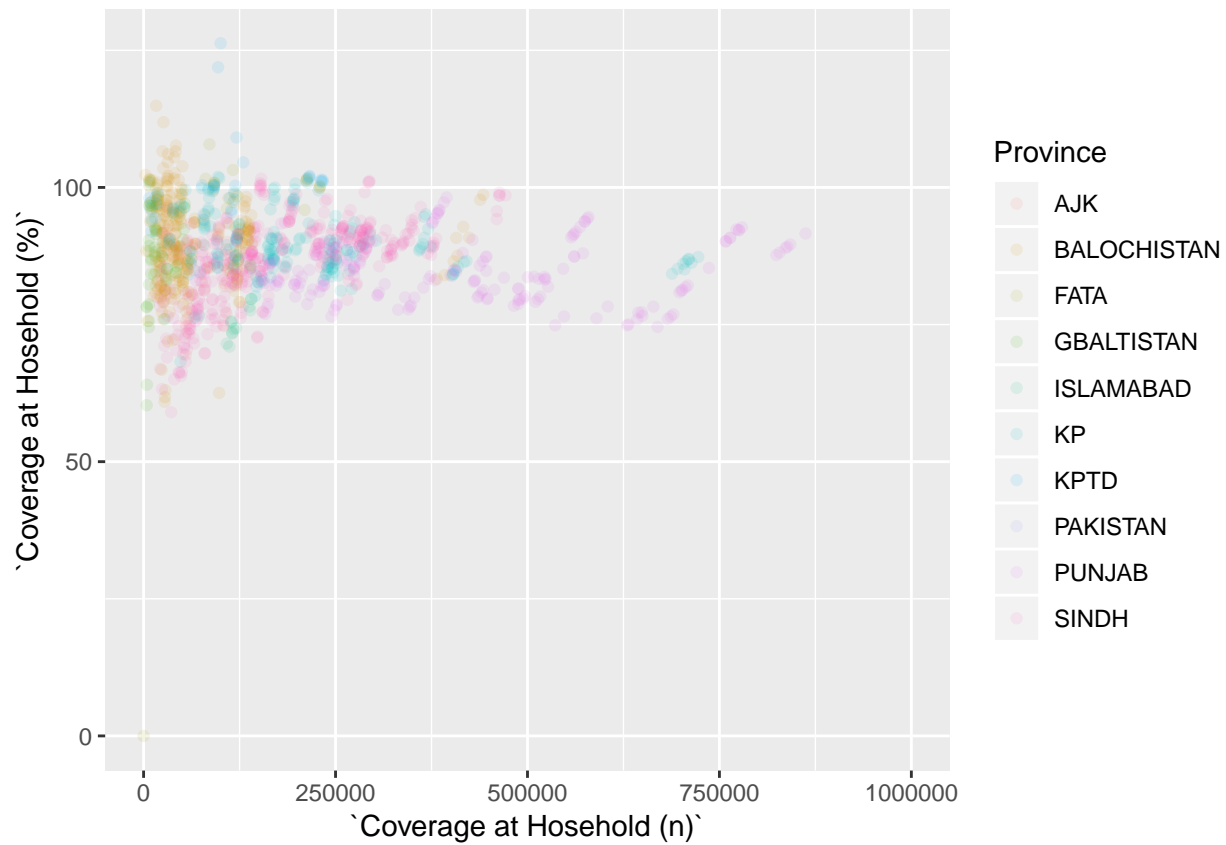
```
my_plot + geom_point(alpha = 0.1, colour = "blue") + xlim(0, 1e+6)
```

Warning: Removed 21 rows containing missing values (geom_point).
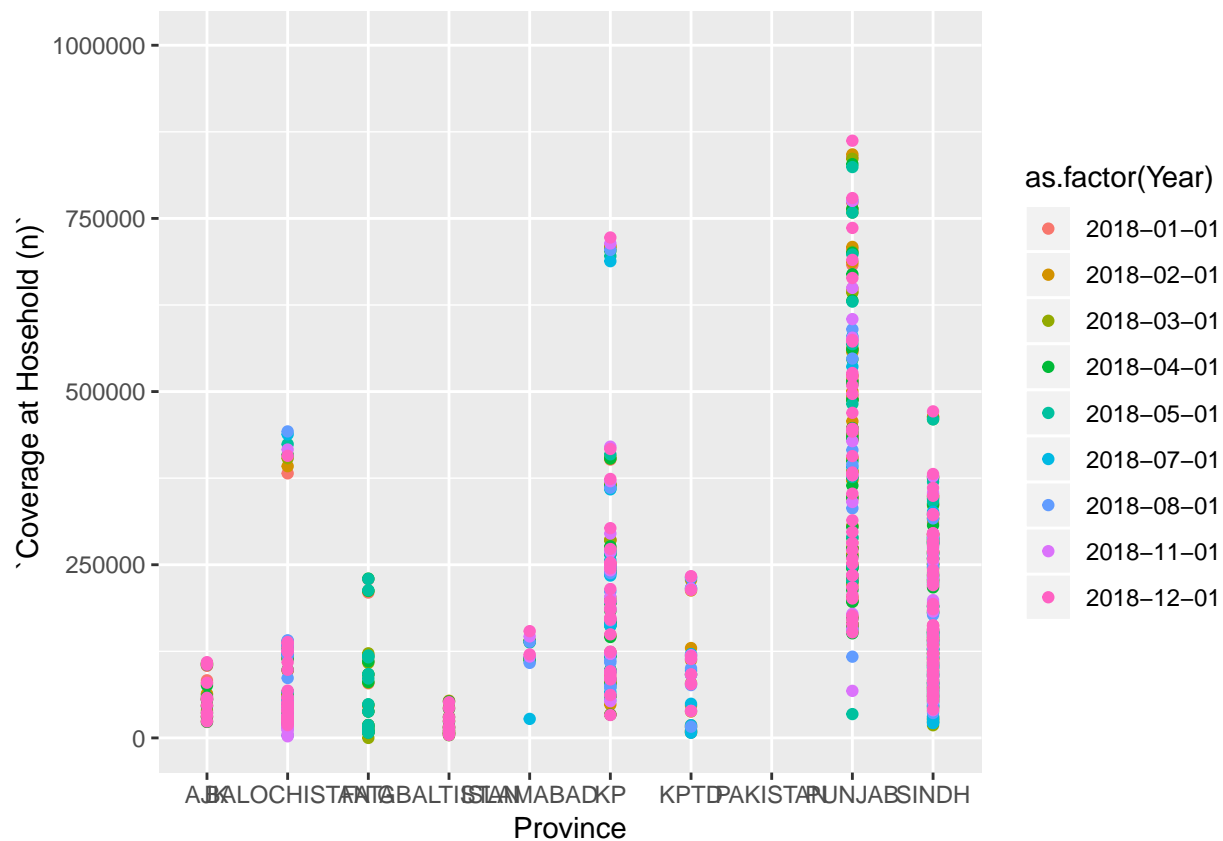
```
my_plot + geom_point(alpha = 0.1, aes(colour = Province)) + xlim(0, 1e+6)
```

```
## Warning: Removed 21 rows containing missing values (geom_point).
```
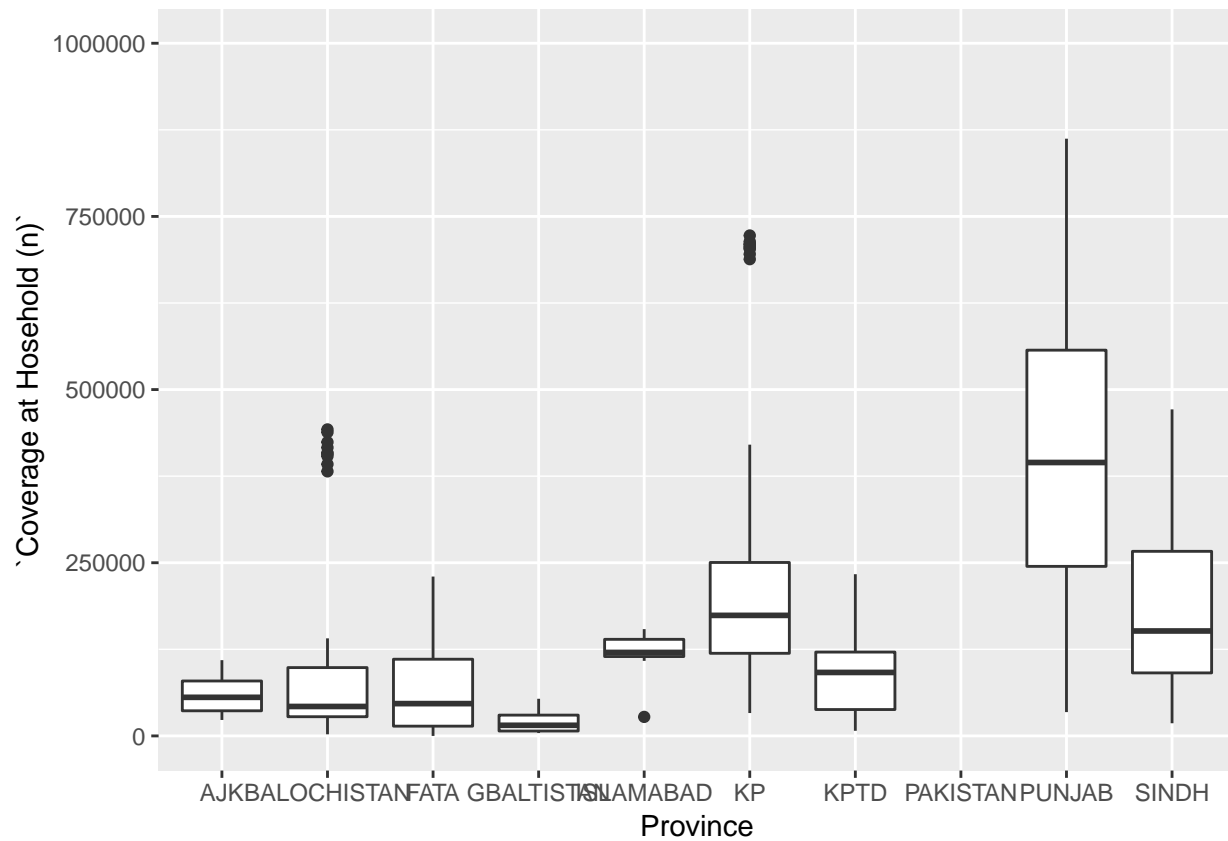
```
ggplot(data = dat,
       mapping = aes(x = Province, y = `Coverage at Hosehold (n)`)) +
  geom_point(aes(colour = as.factor(Year))) + ylim(0, 1e+6)
```

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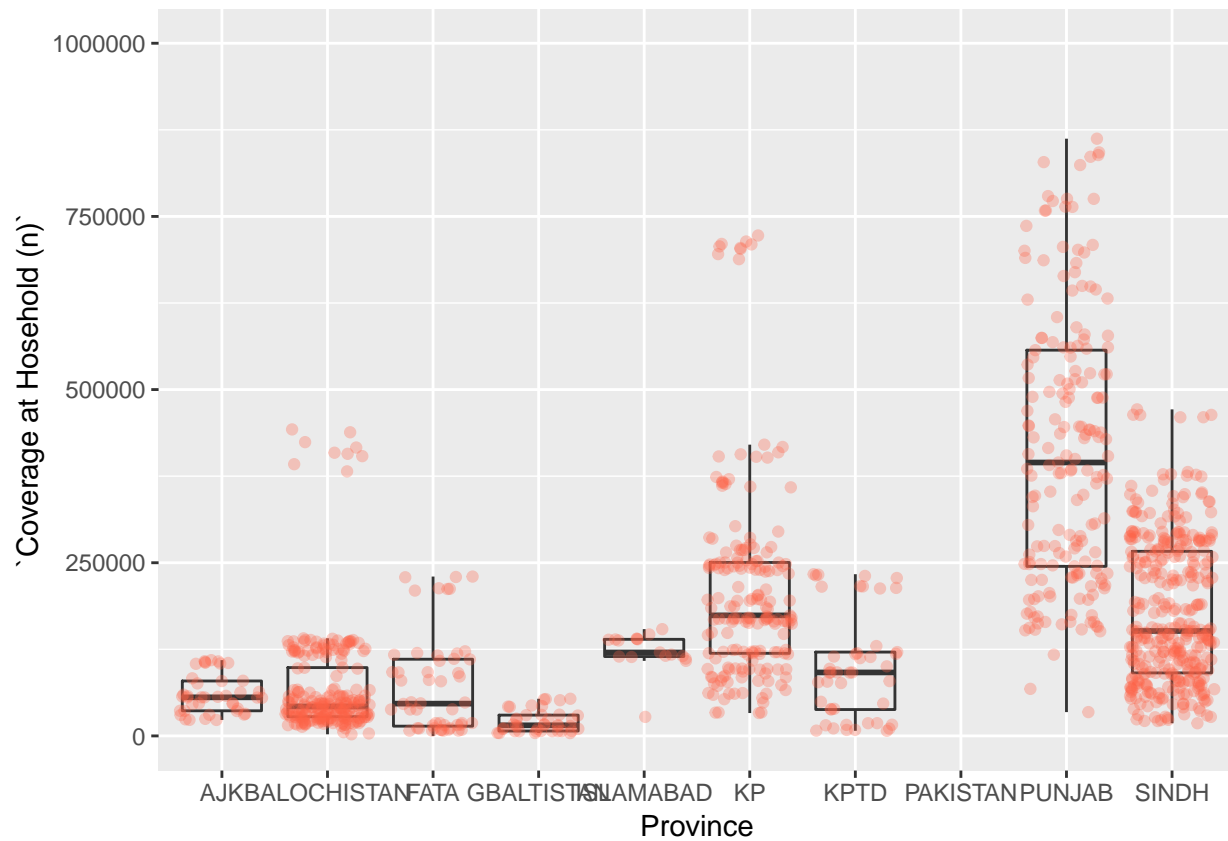
Warning: Removed 21 rows containing non-finite values (stat_boxplot).

```
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 22 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)
```
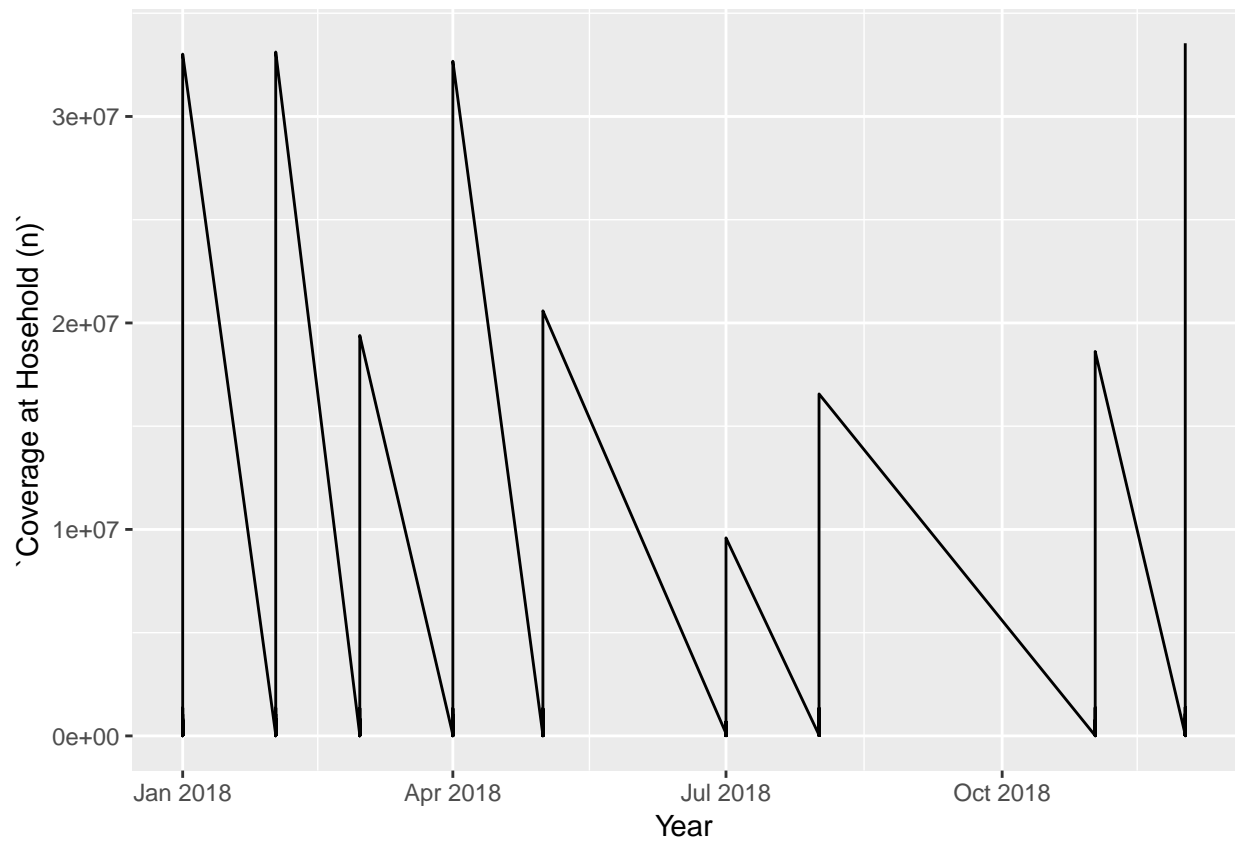
Warning: Removed 21 rows containing non-finite values (stat_boxplot).

Warning: Removed 21 rows containing missing values (geom_point).
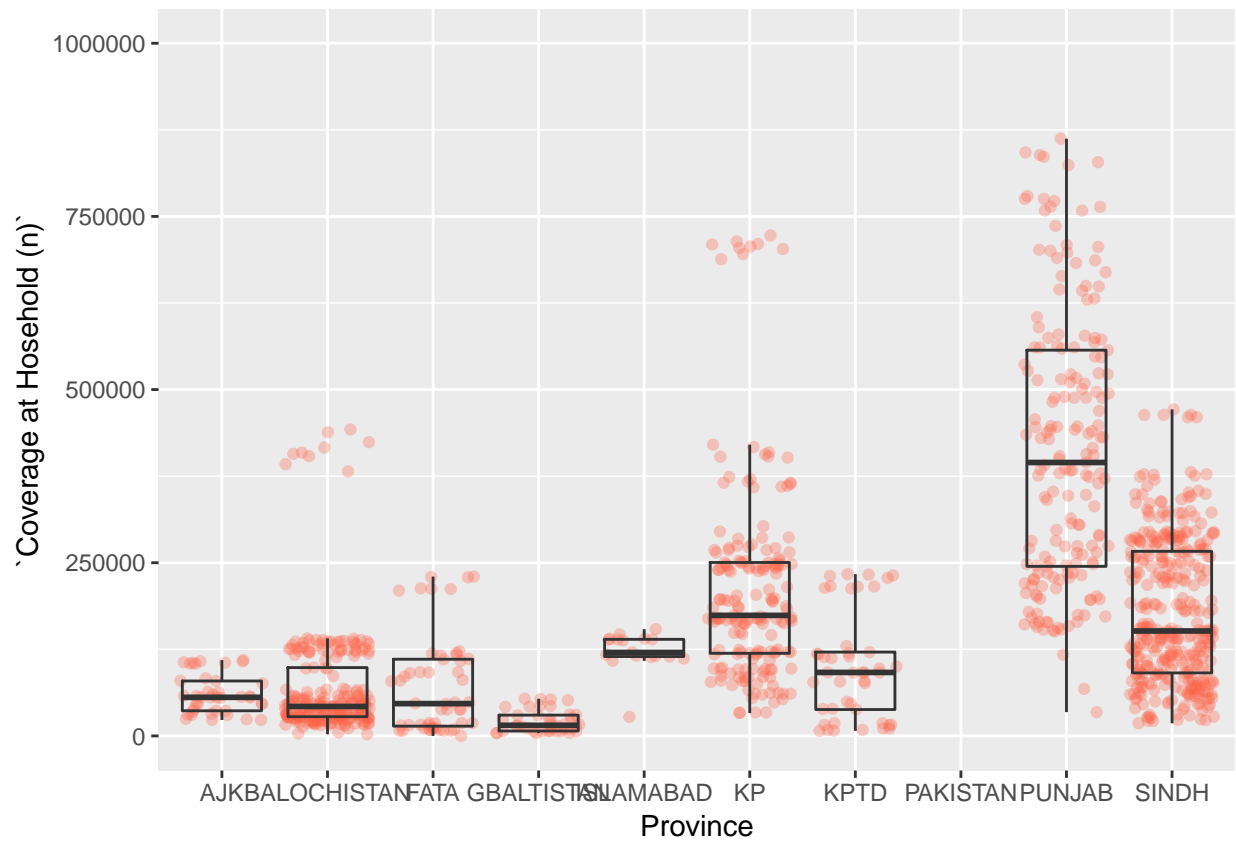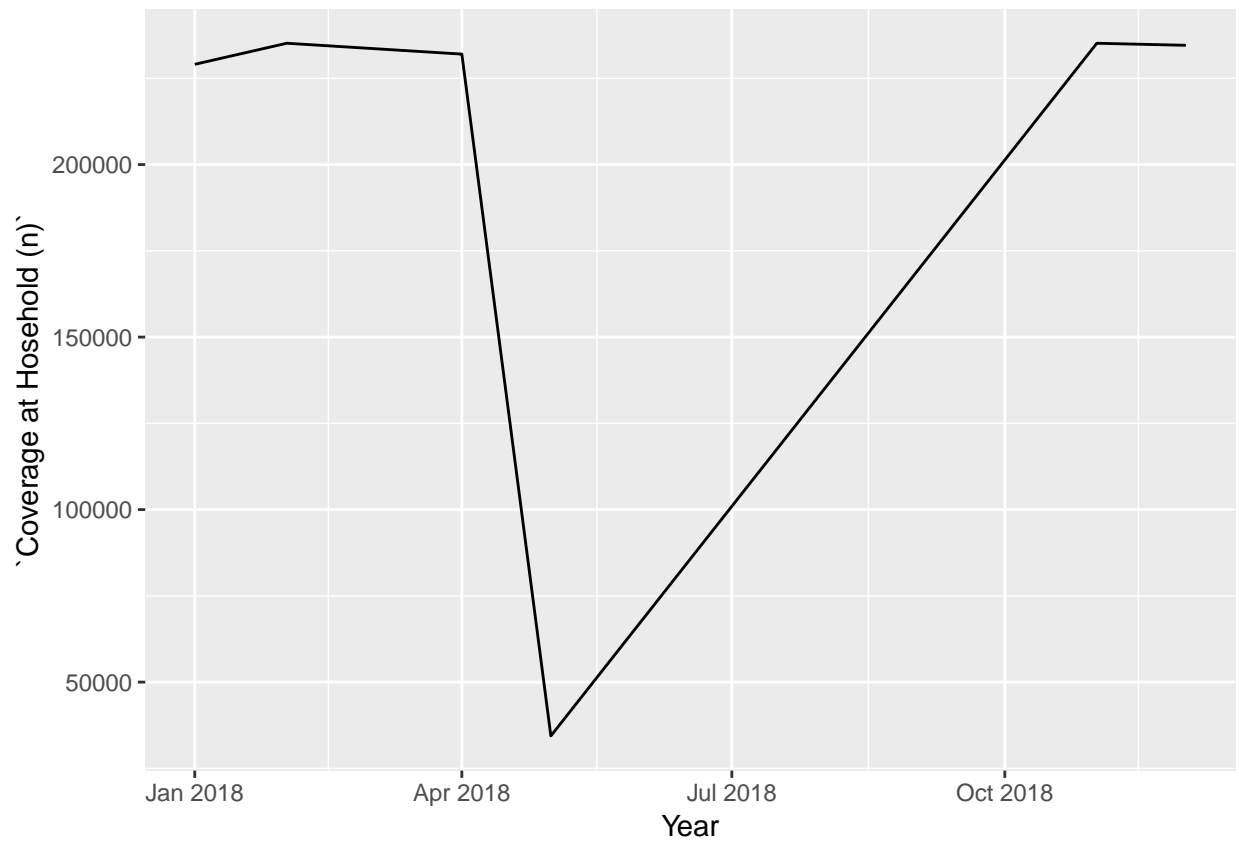
time series

```
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_lir
```

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

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

GHIZER | JHANG | KHIKAMARI | KSAIFULAH | MULTAN
GHOTKI | JHELUM | KHIKORANGI | KURRAM | MUSAKHI
GILGIT | KABDULAH | KHILANDHI | LAHORE | MUZAFFA
GUJRANWALA | KALAT | KHILAYARI | LAKKIMRWT | MUZFARG
GUJRAT | KAMBAR | KHILIAQAT | LARKANA | NAGAR
GWADUR | KARAK | KHIMALIR | LASBELA | NANKANA
HAFIZABAD | KASHMORE | KHINNAZIM | LAYYAH | NAROWA
HANGU | KASUR | KHINORTH | LODHRAN | NEELUM
HARIPUR | KECH | KHIORANGI | LORALAI | NFEROZ
HARNAI | KHAIRPUR | KHISADDAR | MALAKAND | NOSHKI
HAVELI | KHANEWAL | KHISHAHFAISAL | MANSEHRA | NOWSHE
HUNZA | KHARAN | KHISITE | MARDAN | NSIRABA
HYDERABAD | KHARMANG | KHUSHAB | MASTUNG | OKARA
ICT | KHIBALDIA | KHUZDAR | MATIARI | ORAKZAI
JACOBABAD | KHIBINQASIM | KHYBER | MBDIN | PAKPATTE
JAFARABAD | KHIGADAP | KOHAT | MIANWALI | PANJGOU
JAMSHORO | KHIGIQBAL | KOHISTAN | MIRPUR | PESHAW
JEHLUM_VALLEY | KHIGULBERG | KOHLU | MIRPURKHAS | PISHIN
JHALMAGSI | KHIJAMSHEED | KOTLI | MOHMAND | POONCH
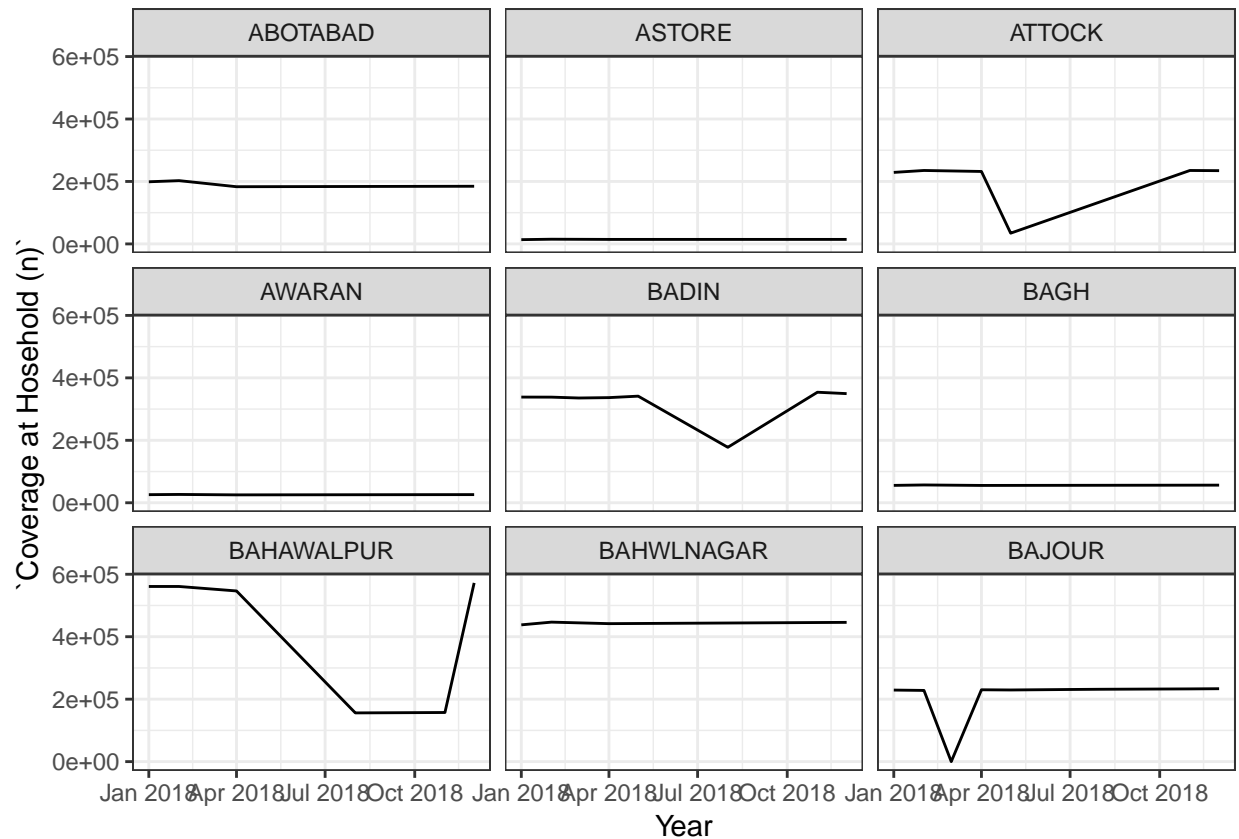
faceting

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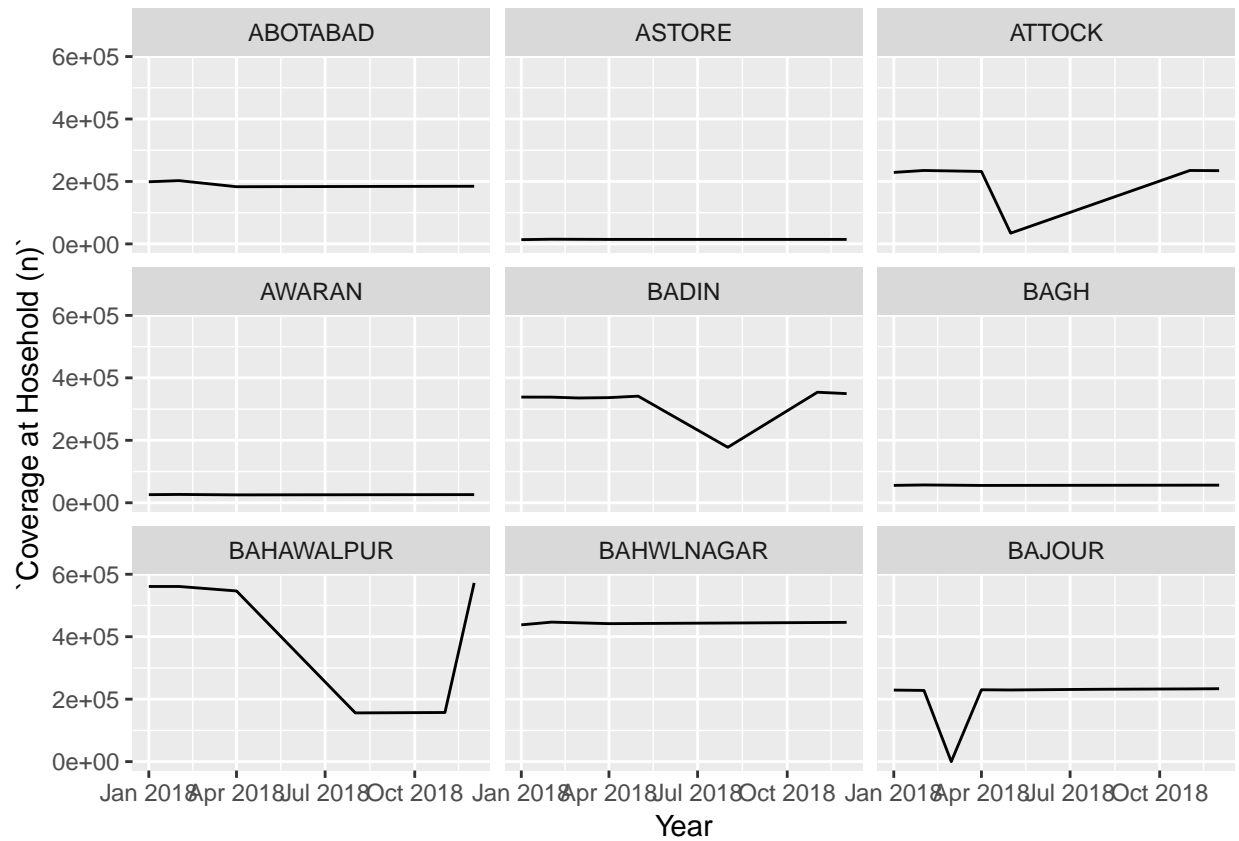
```
ggplot(data = ss_dat, mapping = aes(x = Year, y = `Coverage at Hosehold (n)`)) +
  geom_line() +
  facet_wrap(facets = vars(District)) +
  theme_bw()
```

arranging and exporting plots

```r
# install.packages("gridExtra")
library(gridExtra)
```
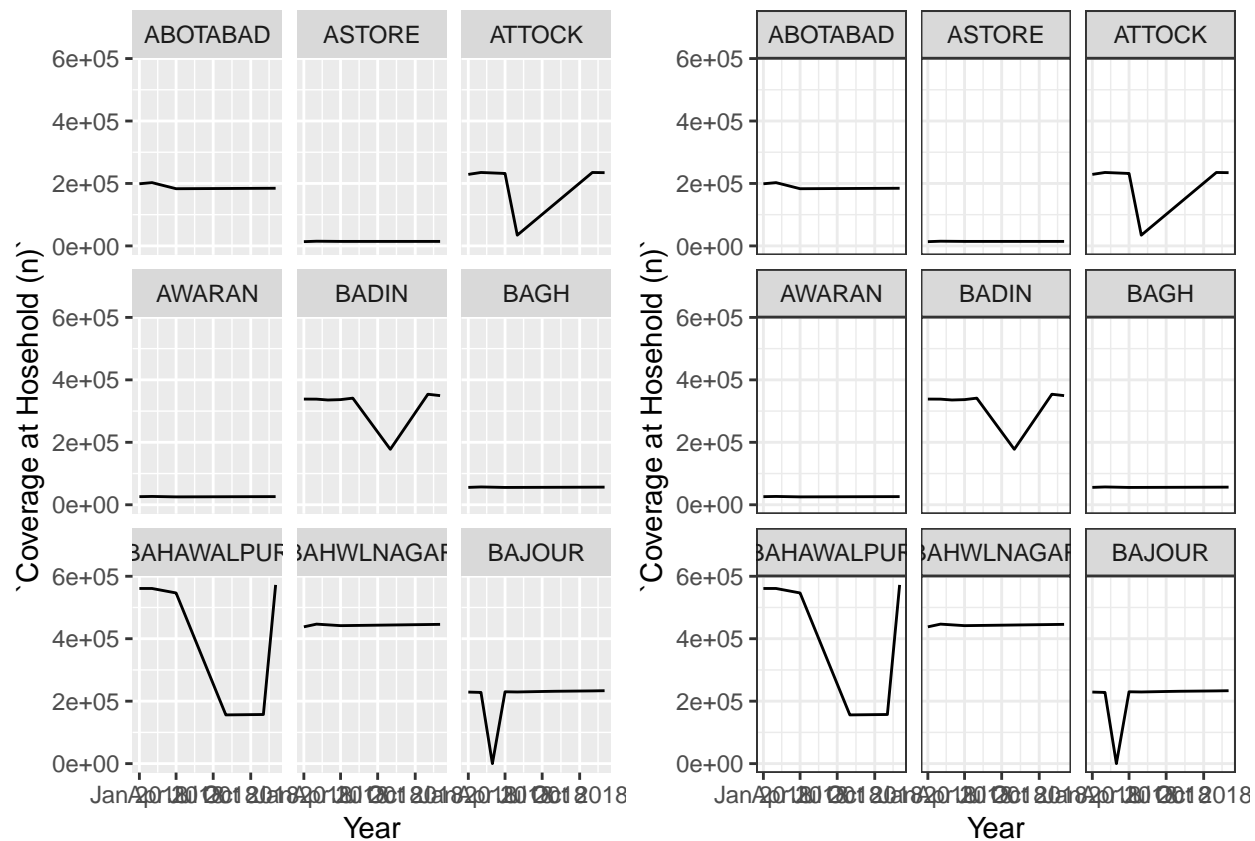
```
##
## Attaching package: 'gridExtra'

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

```r
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)
```

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