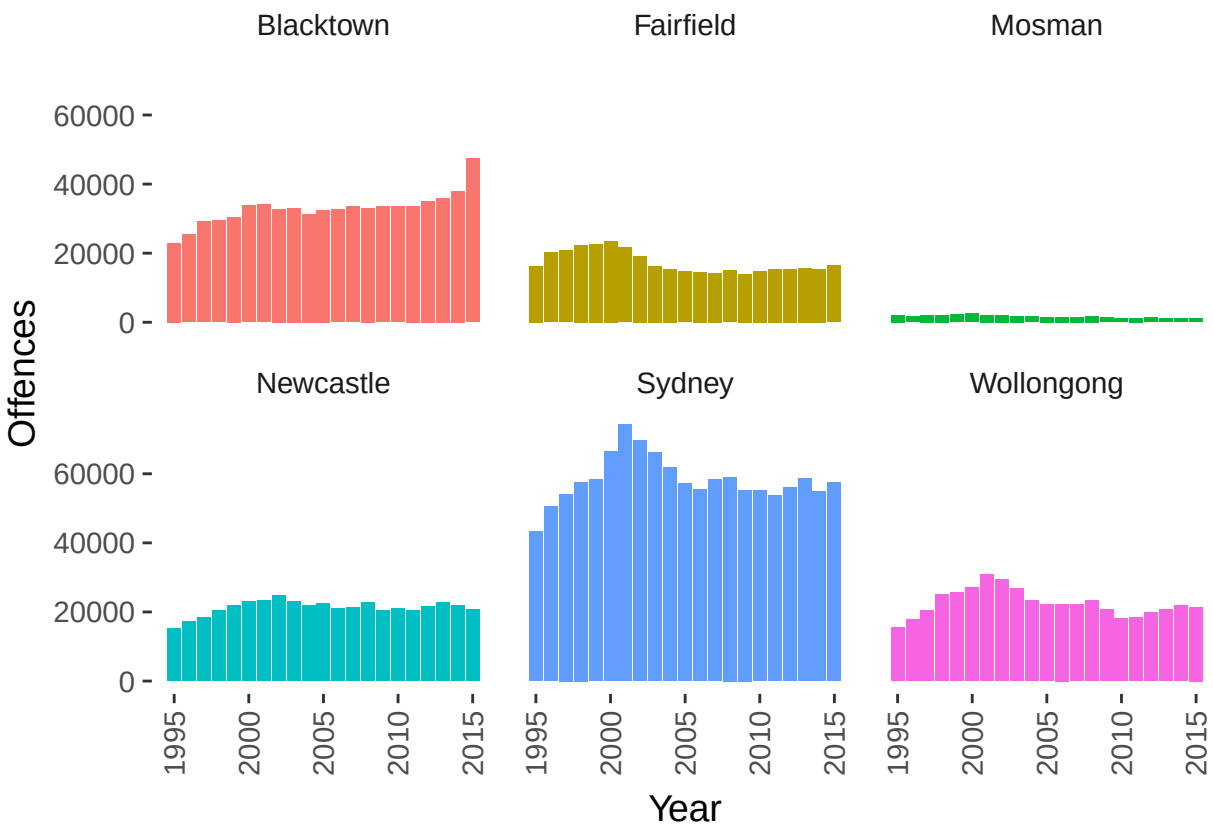
Police Resourcing

Start by importing the offense data from Tim's Excel workbook. To help with visualising the data we need to pivot to narrow format, fix up the dates, and tidy up the columns a bit.

```
library(dplyr)
library(tidyr)
library(readxl)
library(lubridate)
offence_by_month <-
  read_excel("~/Desktop/Tim/RCI_offencebymonth.xlsm") %>%
  gather(key = "Date", value = "Count", -LGA, -`Offence category`, -Subcategory) %>%
  mutate(Date = as.Date(as.numeric(Date), origin = "1900-01-01")) %>%
  mutate(Date = make_date(year(Date), month(Date))) %>% # Making sure dates are 1st of month
  rename(Offence = `Offence category`) %>%
  select(LGA, Date, Offence, Subcategory, Count)
```

We can now plot the number of offences per month over time for the 5 LGAs we're interested in, plus Wollongong (mainly because it makes the plot look better to have 6 instead of 5).

```
library(ggplot2)
library(ggthemes)
offence_by_month %>%
  filter(LGA %in% c("Sydney", "Blacktown", "Newcastle", "Fairfield", "Mosman", "Wollongong")) %>%
  ggplot(aes(x=year(Date), weight=Count, fill=LGA)) +
    geom_bar() +
    facet_wrap(~LGA) +
    xlab("Year") + ylab("Offences") +
    theme_tufte(base_size = 14, base_family = "sans") +
    theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5), legend.position = "none")
```



So that all works. Cool.

Now we need to go and get the police strength (staffing) figures. The datasets are a bit of a mess, with a mix of HTML pages and PDFs, and with a constantly changing web address. I've manually worked through all of these issues and kept everything in an Excel sheet to make it a bit easier to manage.

This takes a few minutes to run.

```
library(rvest)
library(stringr)
library(readr)
library(lubridate)
library(pdftools)

# This is a very messy set of data with a constantly changing format. The script
# below works at the moment, but it's not pretty, and there are still a few
# niggling issues. But it's good enough for now!

controls <- read_excel("~/code/police/scraping_controls.xlsx", na = "NA")
controls <- controls[!is.na(controls$URL),]

monthly_data <- vector(mode = "list", length = nrow(controls))

for (i in 1:nrow(controls)) {

  month_name <- format(make_date(month=controls$Month[i]), "%B")
  year <- controls$Year[i]
  message("Extracting data for ", month_name, " ", year)
```

```
  if (controls$Format[i] == "html") {
    dat <-
      html(controls$URL[i]) %>%
      html_node("#content-main div") %>%
      html_text() %>%
      str_replace_all("\r\n\r\n", "|") %>%
      str_replace_all("\r\n", "\n") %>%
      read_delim(delim = "|")
    names(dat) <- c("REGION", "LAC", "AUTH", "ACTUALS")
    keep_up_to <- which(str_detect(dat$REGION, "Command"))[1] - 1
    # Drop first row and all crud at the end, also drop first column
    dat <- dat[2:keep_up_to,-1]
    # Drop totals
    total_rows <- which(str_detect(dat$LAC, "Total") | str_detect(dat$LAC, "^\\s*$"))
    dat <- dat[-total_rows,]
    dat$DATE <- make_date(controls$Year[i], controls$Month[i])
    dat$LAC <- str_replace(dat$LAC, "^ *", "")
    monthly_data[[i]] <- dat

  } else if (controls$Format[i] == "pdf") {

    dat <- pdf_text(controls$URL[i])[1] %>% str_split("\n")
    header_row <- which(str_detect(dat[[1]], "Region.*Auth.*Actual"))
    dat <- dat[[1]][-c(1:header_row)]
    dat <- dat[-which(dat == "" | str_detect(dat, "Use Only"))]
    total_rows <- which(str_detect(dat, "Total"))
    dat <- dat[-total_rows]
    dat <- str_split(dat, "  +")
    dat <- as.data.frame(t(data.frame(dat)))
    row.names(dat) <- 1:nrow(dat)
    names(dat) <- c("REGION", "LAC", "AUTH", "ACTUALS")
    dat$DATE <- make_date(controls$Year[i], controls$Month[i])
    dat$LAC <- str_replace(dat$LAC, "^ *", "")
    monthly_data[[i]] <- dat[,-1]

  }
}
police_strength <- do.call("rbind", monthly_data)
police_strength$AUTH <- as.numeric(as.character(police_strength$AUTH))
police_strength$ACTUALS <- as.numeric(as.character(police_strength$ACTUALS))
police_strength <- police_strength[!is.na(police_strength$LAC),]
police_strength[police_strength$LAC == "Central Metropolitan","LAC"] <- "Central Metro"
police_strength[str_detect(police_strength$LAC, "Sydney City"),"LAC"] <- "Sydney City"
```

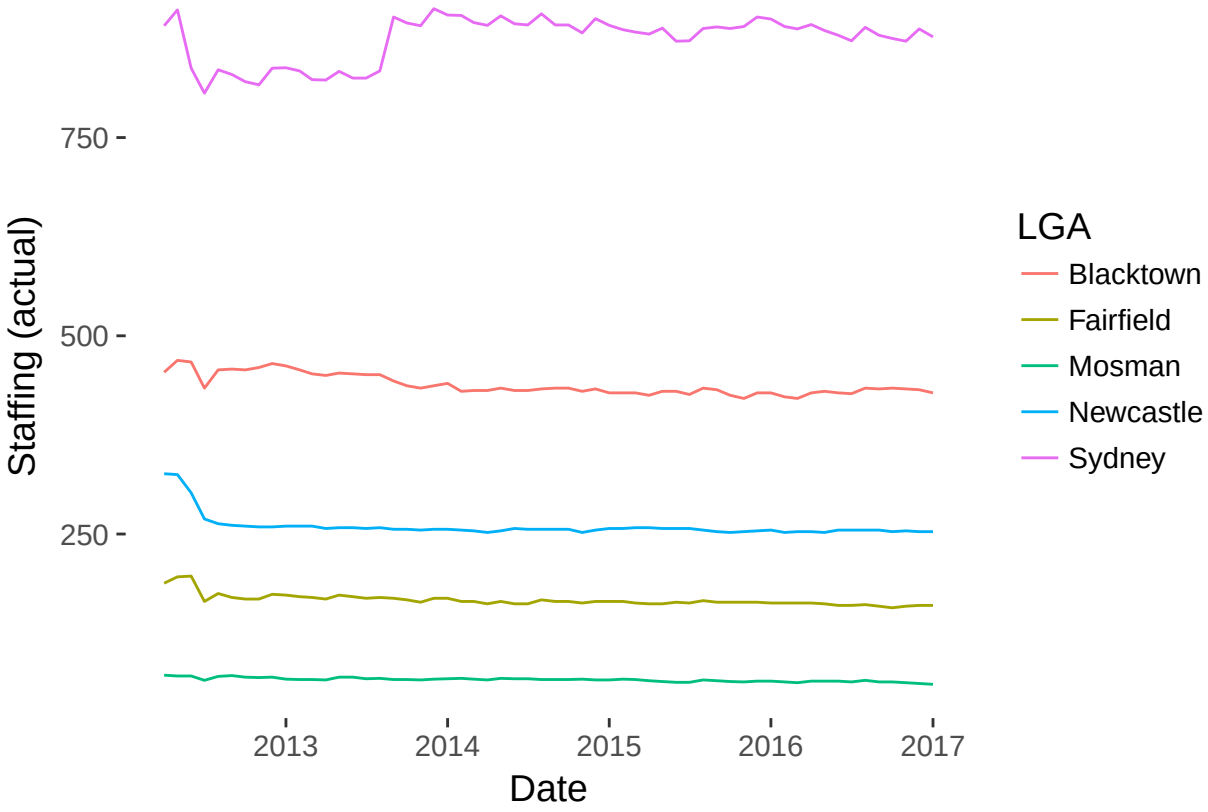
With all of the data collected, we now need to filter and process so it's ready for plotting. This requires a bunch of data cleaning, then a filter and a pivot, followed by the creation of some new columns for the LGA staffing levels (based on the formula given by Tim). We also need to manually insert the value for Fairfield LGA for July 2013, as it somehow got missed and it's quicker to fix manually than it is to fix the import process.

```
# Fixing most of the NAs
missing_rows <- which(is.na(police_strength$ACTUALS))
police_strength$ACTUALS[missing_rows] <- 0
```

```
library(tidyr)
police_strength_LGA <-
police_strength %>%
  # Fixing some data quality issues with LAC names
  mutate(LAC = str_replace(LAC, "(Quakers Hill.*)", "Quakers Hill")) %>%
  mutate(LAC = str_replace(LAC, "Bktown", "Blacktown")) %>%
  # Fixing a name change
  mutate(LAC = str_replace(LAC, "City Central", "Sydney City")) %>%
  # Filter for just the stations we care about
  filter(LAC %in% c("Sydney City", "Central Metro", "City Central",
                    "Kings Cross", "Redfern", "Leichhardt", "Surry Hills",
                    "Blacktown", "Quakers Hill", "Mt Druitt", "Newcastle City",
                    "Fairfield", "Harbourside")) %>%
  # Pivot into a wide table
  select(-AUTH) %>%
  spread(LAC, ACTUALS) %>%
  # Calulate LGA staffing using Tim's formula
  mutate(SydneyLGA =
           `Sydney City` +
           `Central Metro` +
           `Kings Cross` +
           `Redfern` +
           `Surry Hills` +
           (`Leichhardt`/2),
         BlacktownLGA =
           `Blacktown` +
           `Quakers Hill` +
           `Mt Druitt`,
         NewcastleLGA =
           `Newcastle City`,
         FairfieldLGA =
           `Fairfield`,
         MosmanLGA =
           (`Harbourside`/2)) %>%
  # Select just the LGAs (remove all of the LACs)
  select(DATE, SydneyLGA, BlacktownLGA, NewcastleLGA, FairfieldLGA, MosmanLGA)

  # Somehow lost a single value for Fairfield, so inserting it manually
  fix_row <- which(police_strength_LGA$DATE == "2013-07-01")
  police_strength_LGA[fix_row, "FairfieldLGA"] <- 169
```

Cool. Now we can plot how each LGA's staffing changed over time.

```
police_strength_LGA %>%
  # Pivoting back to narrow for plotting
  gather("LGA", "Staff", -DATE) %>%
  # Making things look nicer
  mutate(LGA = str_replace(LGA, "LGA", "")) %>%
  # Plotting
  ggplot(aes(x = DATE, y = Staff, col = LGA)) +
    geom_line() +
    labs(x = "Date", y = "Staffing (actual)") +
    theme_tufte(base_size = 14, base_family = "sans")
```



At this point, we have enough data to try some modelling approaches, however the assumptions for these approaches are going to be difficult to determine, so we can stick to visualisation for now, to see if we can identify if there are any clear stories here.

The code below creates a function which will let us plot each of the LGAs easily.

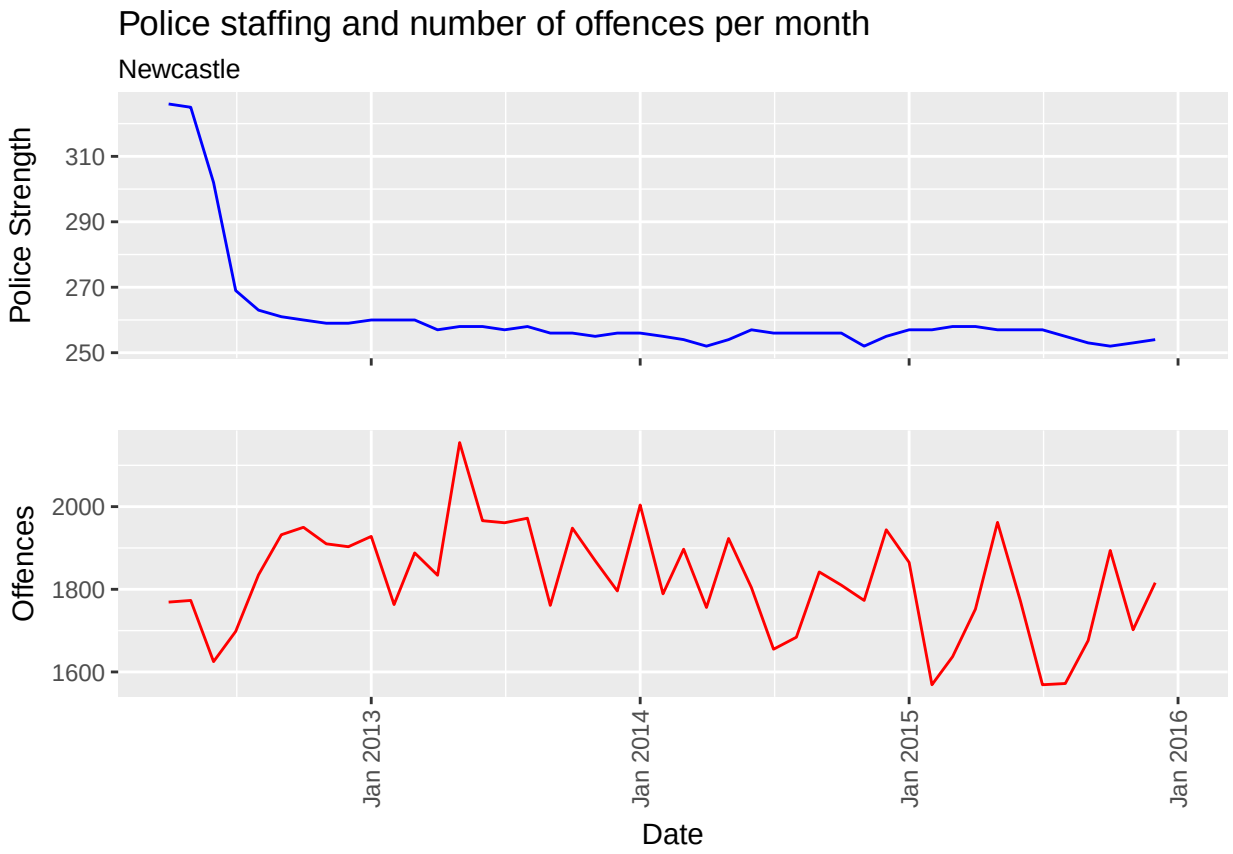
```
library(gtable)
library(grid)

plot_lga <- function(lga_filter) {

  gg1 <-
  police_strength_LGA %>%
  # Pivoting back to narrow for plotting
  gather("LGA", "Staff", -DATE) %>%
  # Making things look nicer
  mutate(LGA = str_replace(LGA, "LGA", "")) %>%
  # Filtering for specified lGA only
  filter(LGA == lga_filter) %>%
  # Plotting
  filter(DATE >= ymd("2012-04-01")) %>%
  filter(DATE <= ymd("2015-12-31")) %>%
  ggplot(aes(x = DATE, y = Staff)) +
    geom_line(col = "blue") +
    labs(x = "", y = "Police Strength", subtitle = lga_filter, title = "Police staffing and number of o
    #theme_tufte(base_size = 14, base_family = "sans") +
    theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5), legend.position = "none") +
    scale_x_date(limits=c(ymd("2012-04-01"), ymd("2015-12-31")), date_breaks = "years", labels = NULL)
```

```
gg2 <-
  offence_by_month %>%
  filter(LGA == lga_filter) %>%
  filter(Date >= "2012-04-01") %>%
  filter(Date <= "2015-12-31") %>%
  group_by(Date, LGA) %>%
  summarise(Count = sum(Count)) %>%
  ggplot() +
  geom_line(aes(x=Date, y=Count), col="red") +
    #theme_tufte(base_size = 14, base_family = "sans") +
    theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5), legend.position = "none") +
    scale_x_date(limits=c(ymd("2012-04-01"), ymd("2015-12-31")), date_breaks = "years", date_labels = "
    labs(x = "Date", y = "Offences")

## convert plots to gtable objects
g1 <- ggplotGrob(gg1)
g2 <- ggplotGrob(gg2)
g <- rbind(g1, g2, size="first") # stack the two plots
g$widths <- unit.pmax(g1$widths, g2$widths) # use the largest widths
grid.newpage()
grid.draw(g)
}
```

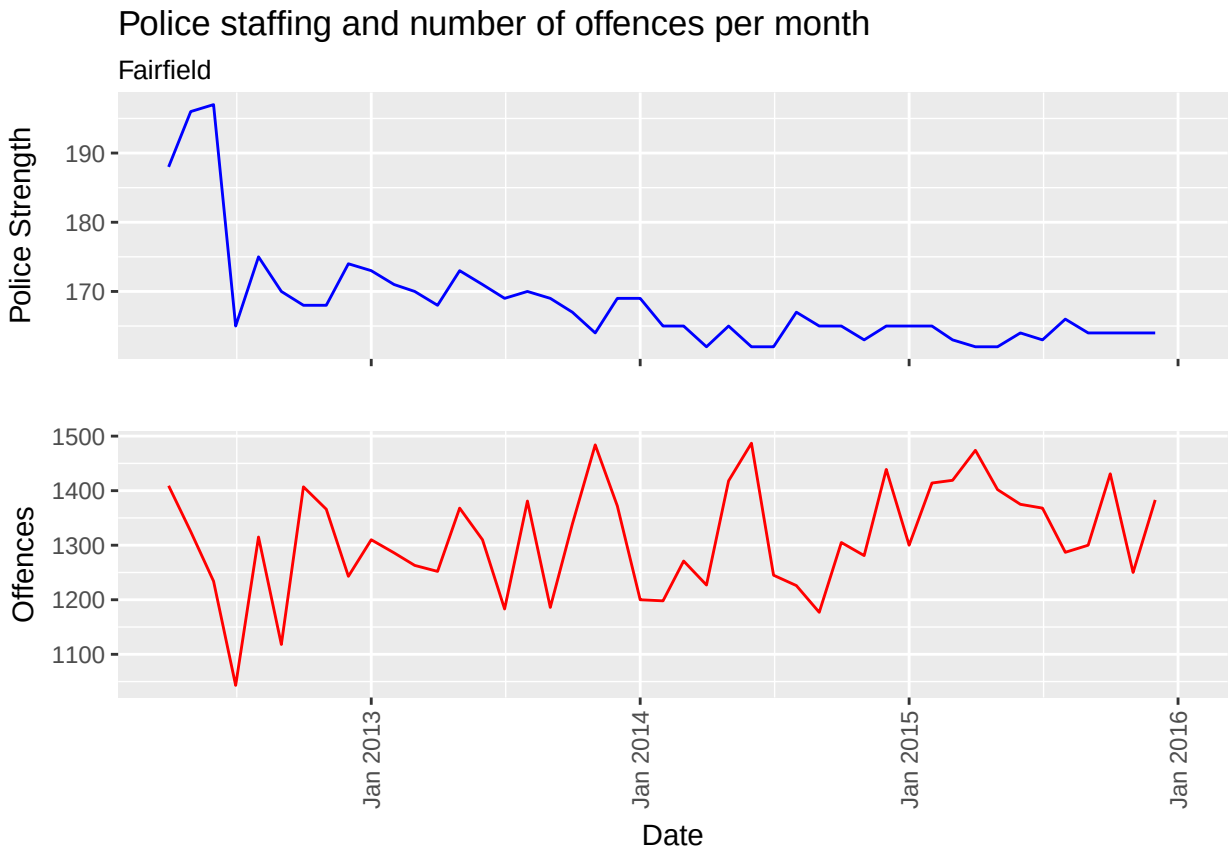
Now to generate some plots...

```
plot_lga("Sydney")
```

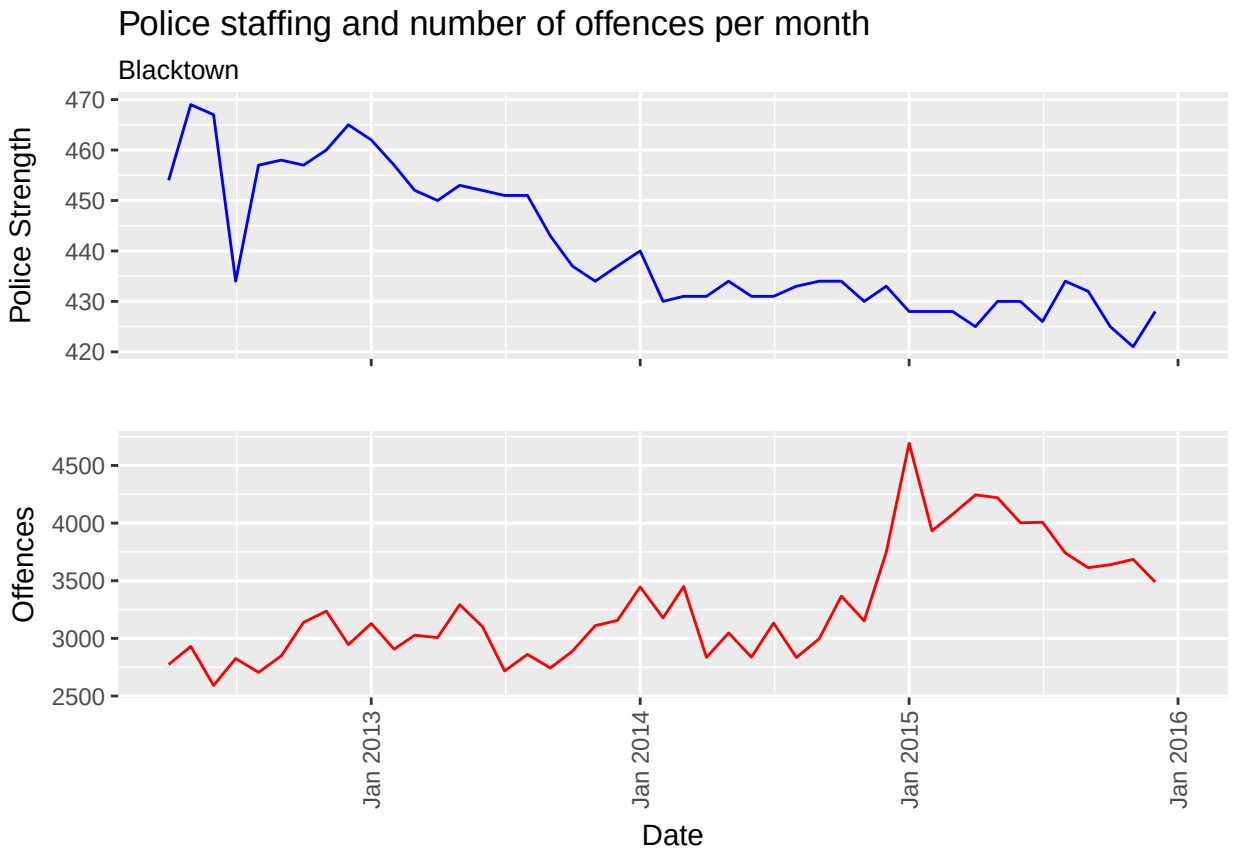
```
plot_lga("Newcastle")
```



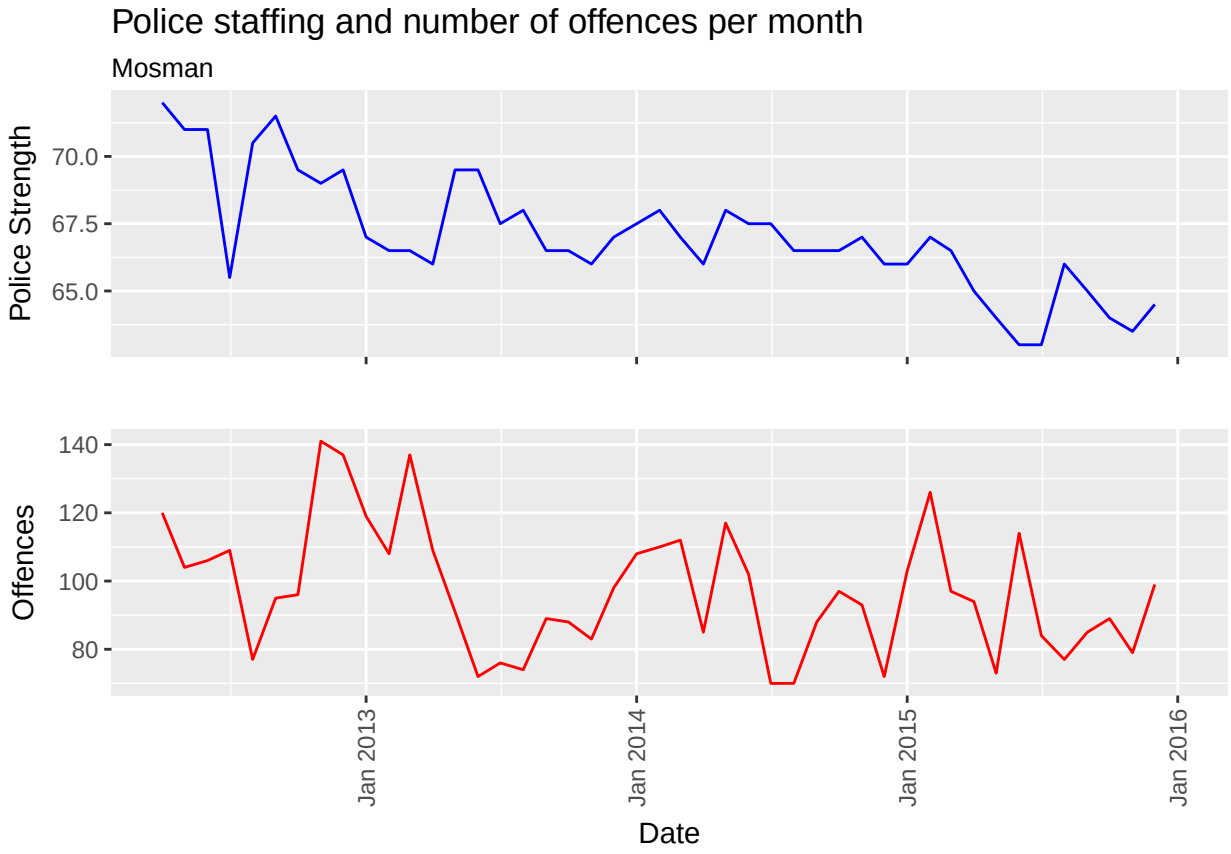

```
plot_lga("Fairfield")
```




```
plot_lga("Blacktown")
```




```
plot_lga("Mosman")
```



One thing we can do to drill down further on this is to group together similar types of crimes. This is going to be a very plot-heavy affair, so we're going to generate them programmatically and make Tim go through them manually. Sucker! Groupings provided by Tim.

```
groupings <- list(
  Homicide = c("Homicide-Murder (a)",
               "Homicide-Attempted murder",
               "Homicide-Murder accessory, conspiracy",
               "Homicide-Manslaughter (a)"),
  `Domestic Assault` = "Assault-Domestic violence related assault",
  Assault = c("Assault-Non-domestic violence related assault",
              "Assault-Assault Police"),
  `Sexual Offences` = c("Sexual offences-Sexual assault",
                        "Sexual offences-Indecent assault, act of indecency and other sexual offences")
  Theft = c("Theft-Break and enter dwelling",
            "Theft-Break and enter non-dwelling",
            "Theft-Receiving or handling stolen goods",
            "Theft-Motor vehicle theft",
            "Theft-Steal from motor vehicle",
            "Theft-Steal from retail store",
            "Theft-Steal from dwelling",
            "Theft-Steal from person",
            "Theft-Stock theft",
            "Theft-Fraud",
            "Theft-Other theft"),
  Robbery = c("Robbery-Robbery without a weapon",
              "Robbery-Robbery with a firearm",
```

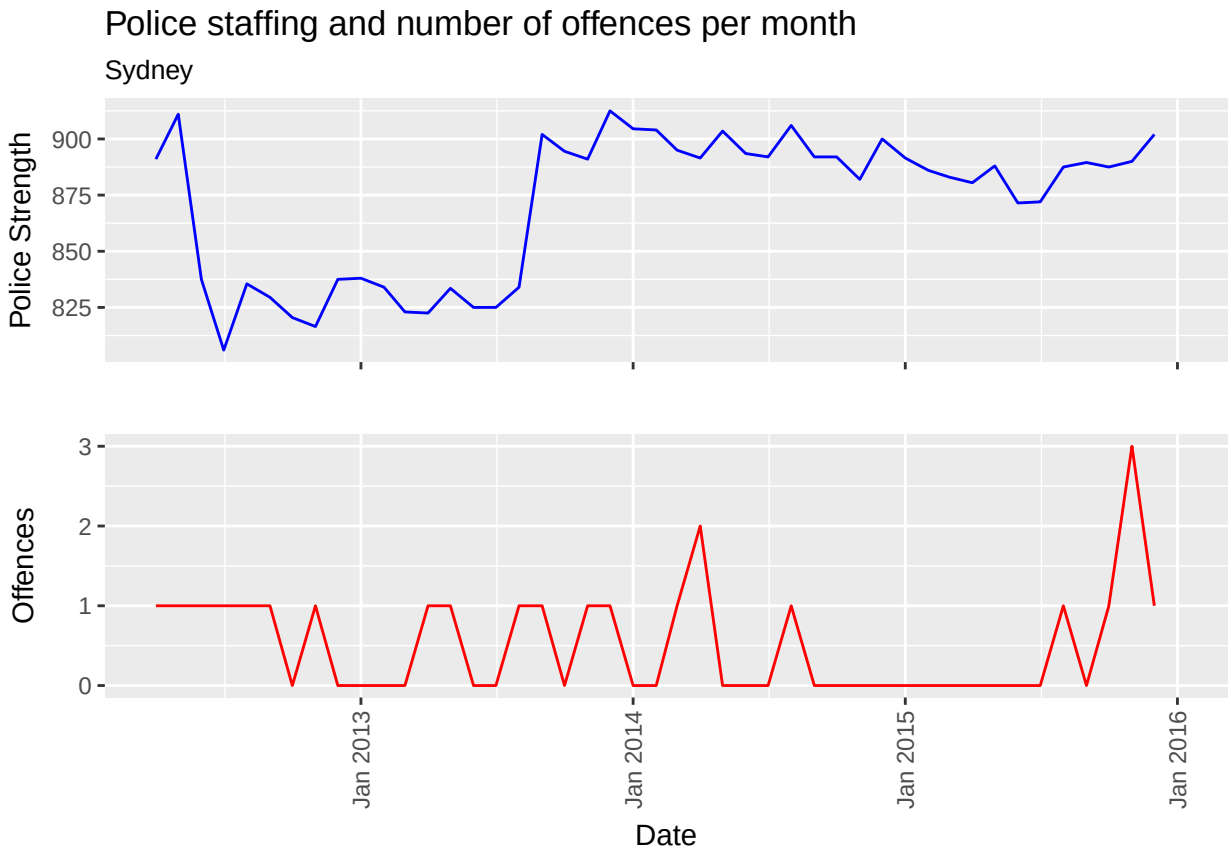
```
              "Robbery-Robbery with a weapon not a firearm"),
  `Drug Possession` = c("Drug offences-Possession and/or use of cocaine",
                        "Drug offences-Possession and/or use of narcotics",
                        "Drug offences-Possession and/or use of cannabis",
                        "Drug offences-Possession and/or use of amphetamines",
                        "Drug offences-Possession and/or use of ecstasy",
                        "Drug offences-Possession and/or use of other drugs"),
  `Drug Dealing` = c("Drug offences-Dealing, trafficking in cocaine",
                     "Drug offences-Dealing, trafficking in narcotics",
                     "Drug offences-Dealing, trafficking in cannabis",
                     "Drug offences-Dealing, trafficking in amphetamines",
                     "Drug offences-Dealing, trafficking in ecstasy",
                     "Drug offences-Dealing, trafficking in other drugs"),
  `Drug Offences` = c("Drug offences-Possession and/or use of cocaine",
                      "Drug offences-Possession and/or use of narcotics",
                      "Drug offences-Possession and/or use of cannabis",
                      "Drug offences-Possession and/or use of amphetamines",
                      "Drug offences-Possession and/or use of ecstasy",
                      "Drug offences-Possession and/or use of other drugs",
                      "Drug offences-Dealing, trafficking in cocaine",
                      "Drug offences-Dealing, trafficking in narcotics",
                      "Drug offences-Dealing, trafficking in cannabis",
                      "Drug offences-Dealing, trafficking in amphetamines",
                      "Drug offences-Dealing, trafficking in ecstasy",
                      "Drug offences-Dealing, trafficking in other drugs",
                      "Drug offences-Cultivating cannabis",
                      "Drug offences-Manufacture drug",
                      "Drug offences-Importing drugs",
                      "Drug offences-Other drug offences"),
  `Disorderly Conduct` = c("Disorderly conduct-Trespass",
                           "Disorderly conduct-Offensive conduct",
                           "Disorderly conduct-Offensive language",
                           "Disorderly conduct-Criminal intent"),
  `Against Justice Procedures` = c("Against justice procedures-Escape custody",
                                   "Against justice procedures-Breach Apprehended Violence Order",
                                   "Against justice procedures-Breach bail conditions",
                                   "Against justice procedures-Fail to appear",
                                   "Against justice procedures-Resist or hinder officer")
)
```

Also need to extend the plotting function from before...

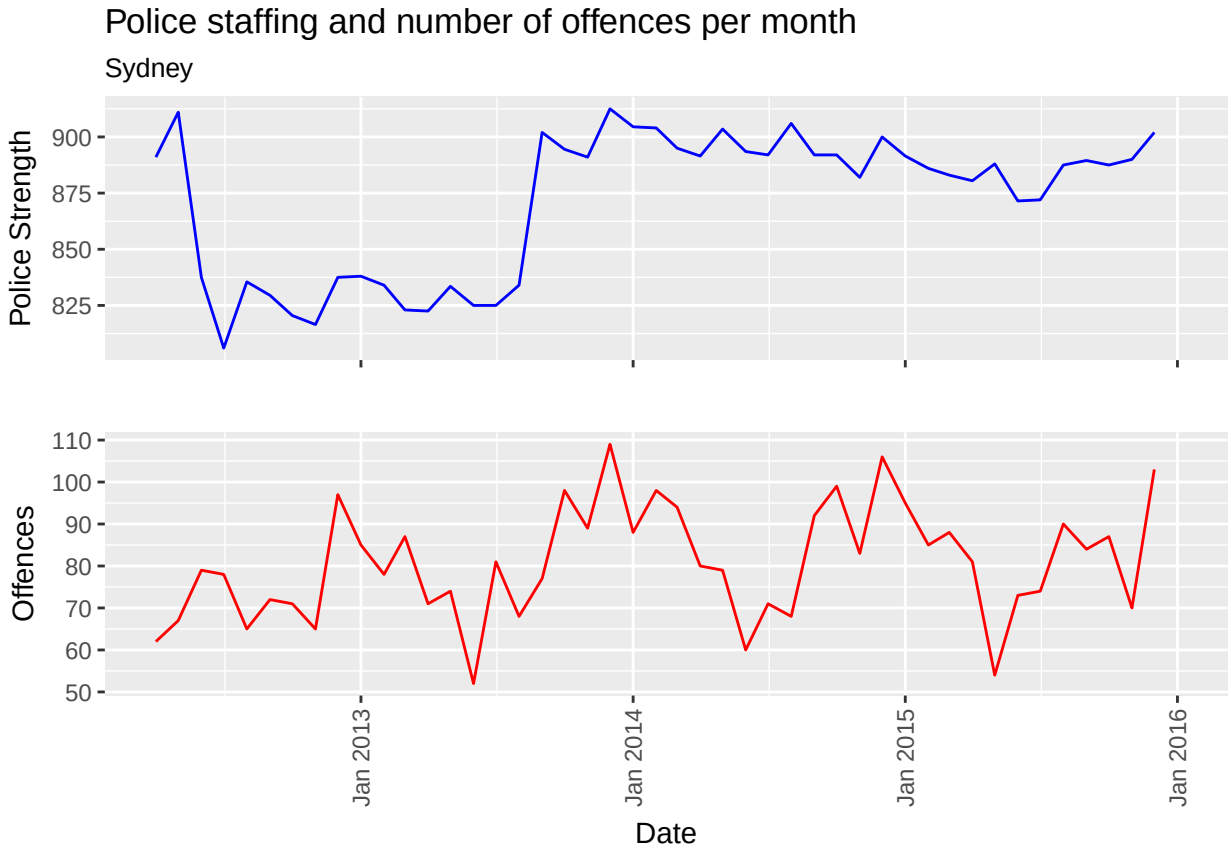
```
plot_lga_filtered <- function(lga_filter, crime_types, group_label) {

  gg1 <-
  police_strength_LGA %>%
  # Pivoting back to narrow for plotting
  gather("LGA", "Staff", -DATE) %>%
  # Making things look nicer
  mutate(LGA = str_replace(LGA, "LGA", "")) %>%
  # Filtering for specified lGA only
  filter(LGA == lga_filter) %>%
  # Plotting
  filter(DATE >= ymd("2012-04-01")) %>%
```

```
  filter(DATE <= ymd("2015-12-31")) %>%
  ggplot(aes(x = DATE, y = Staff)) +
    geom_line(col = "blue") +
    labs(x = "", y = "Police Strength", subtitle = lga_filter, title = "Police staffing and number of o
    #theme_tufte(base_size = 14, base_family = "sans") +
    theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5), legend.position = "none") +
    scale_x_date(limits=c(ymd("2012-04-01"), ymd("2015-12-31")), date_breaks = "years", labels = NULL)

gg2 <-
  offence_by_month %>%
  filter(LGA == lga_filter) %>%
  # Filtering for specified crime types only
  filter(paste(Offence, Subcategory, sep = "-") %in% crime_types) %>%
  filter(Date >= "2012-04-01") %>%
  filter(Date <= "2015-12-31") %>%
  group_by(Date, LGA) %>%
  summarise(Count = sum(Count)) %>%
  ggplot() +
  geom_line(aes(x=Date, y=Count), col="red") +
    #theme_tufte(base_size = 14, base_family = "sans") +
    theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5), legend.position = "none") +
    scale_x_date(limits=c(ymd("2012-04-01"), ymd("2015-12-31")), date_breaks = "years", date_labels = "
    labs(x = "Date", y = "Offences")

## convert plots to gtable objects
g1 <- ggplotGrob(gg1)
g2 <- ggplotGrob(gg2)
g <- rbind(g1, g2, size="first") # stack the two plots
g$widths <- unit.pmax(g1$widths, g2$widths) # use the largest widths
grid.newpage()
grid.draw(g)
}
```

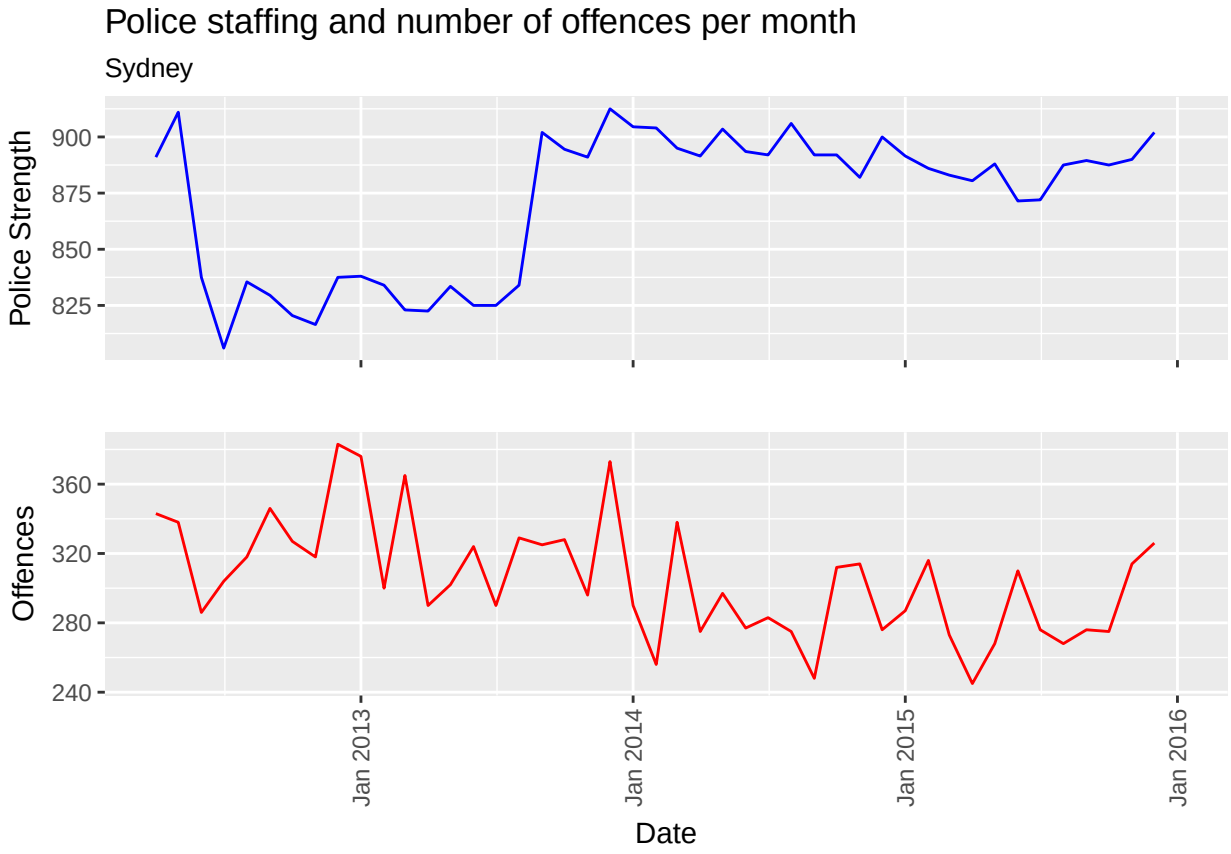
```
LGA_list <- c("Sydney", "Blacktown", "Newcastle", "Fairfield", "Mosman")
for (i in LGA_list) {
  for (j in 1:length(groupings)) {
    plot_lga_filtered(i, groupings[[j]], names(groupings)[j])
    print(paste("LGA: ", i, " Crime type: ", names(groupings)[j]))
  }
}
```



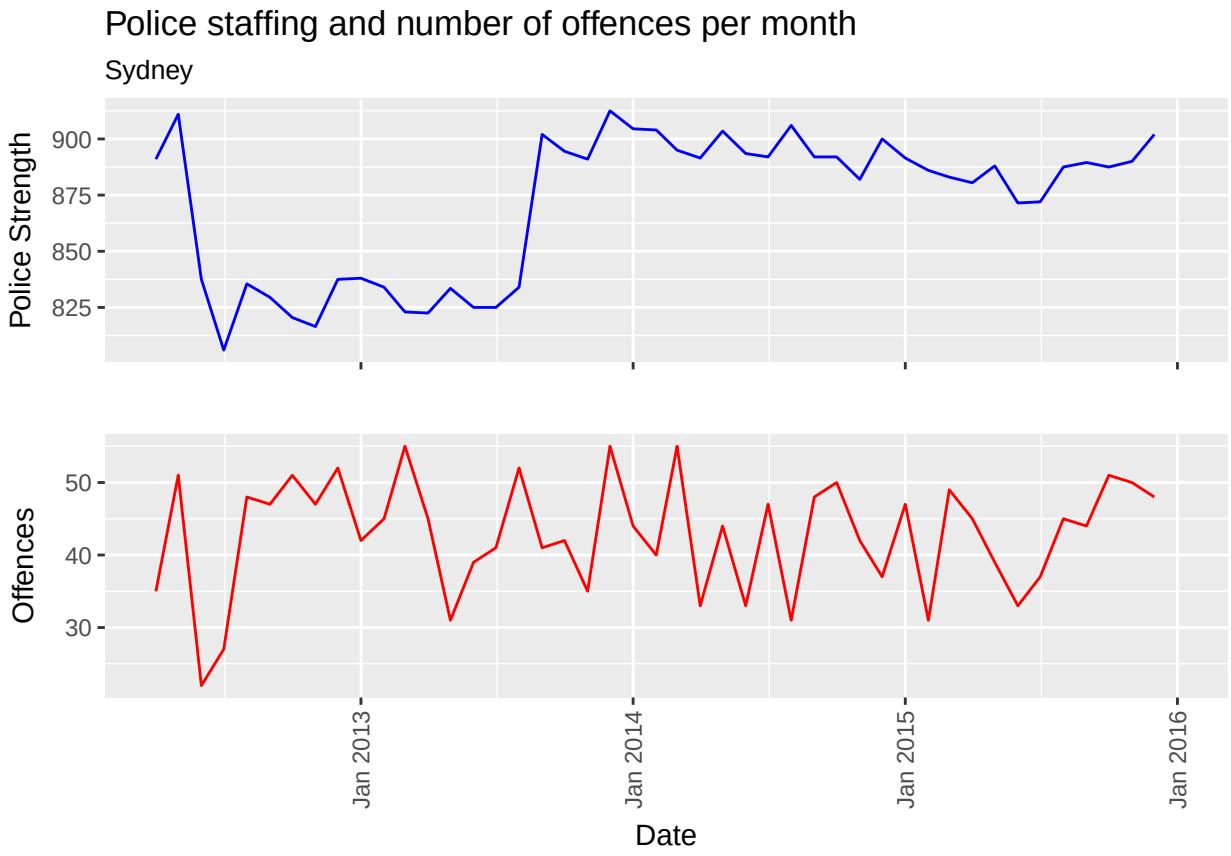

```
## [1] "LGA:  Sydney  Crime type:  Homicide"
```



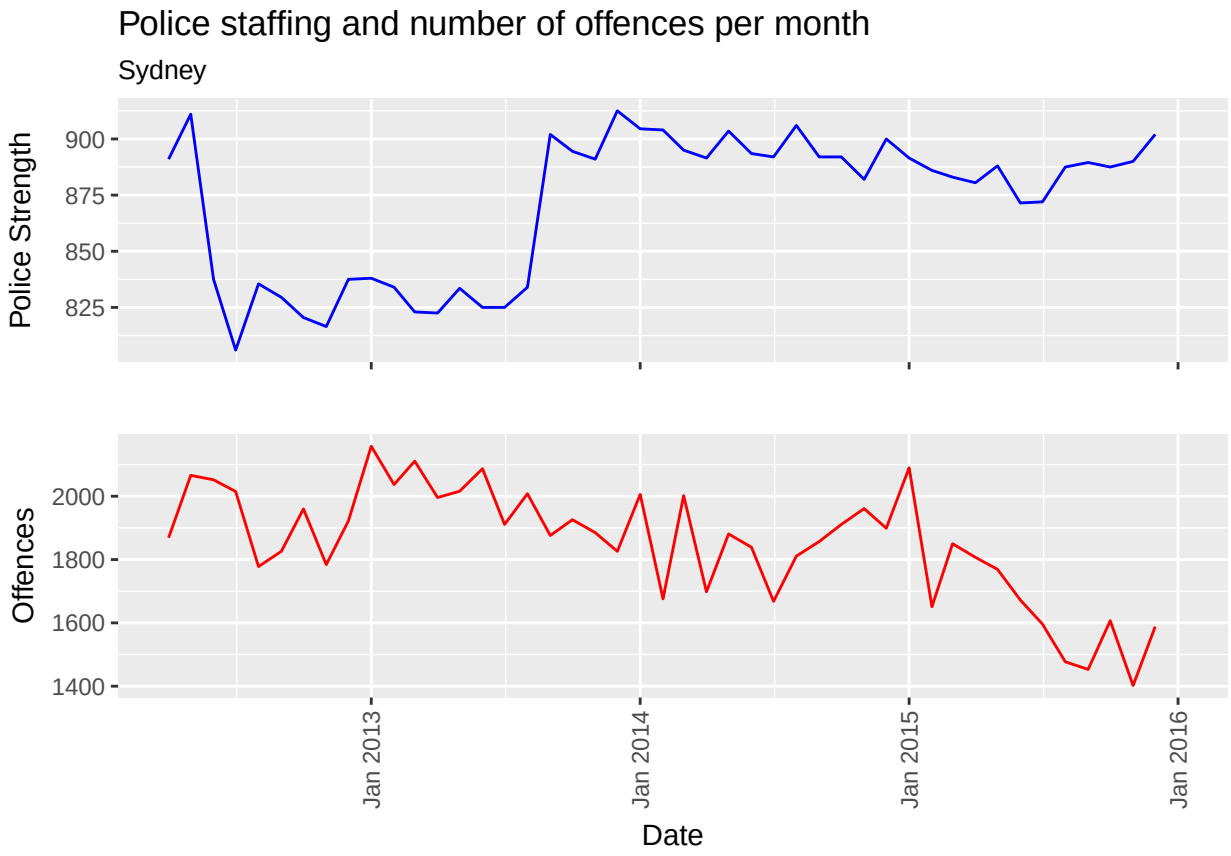

```
## [1] "LGA:  Sydney  Crime type:  Domestic Assault"
```



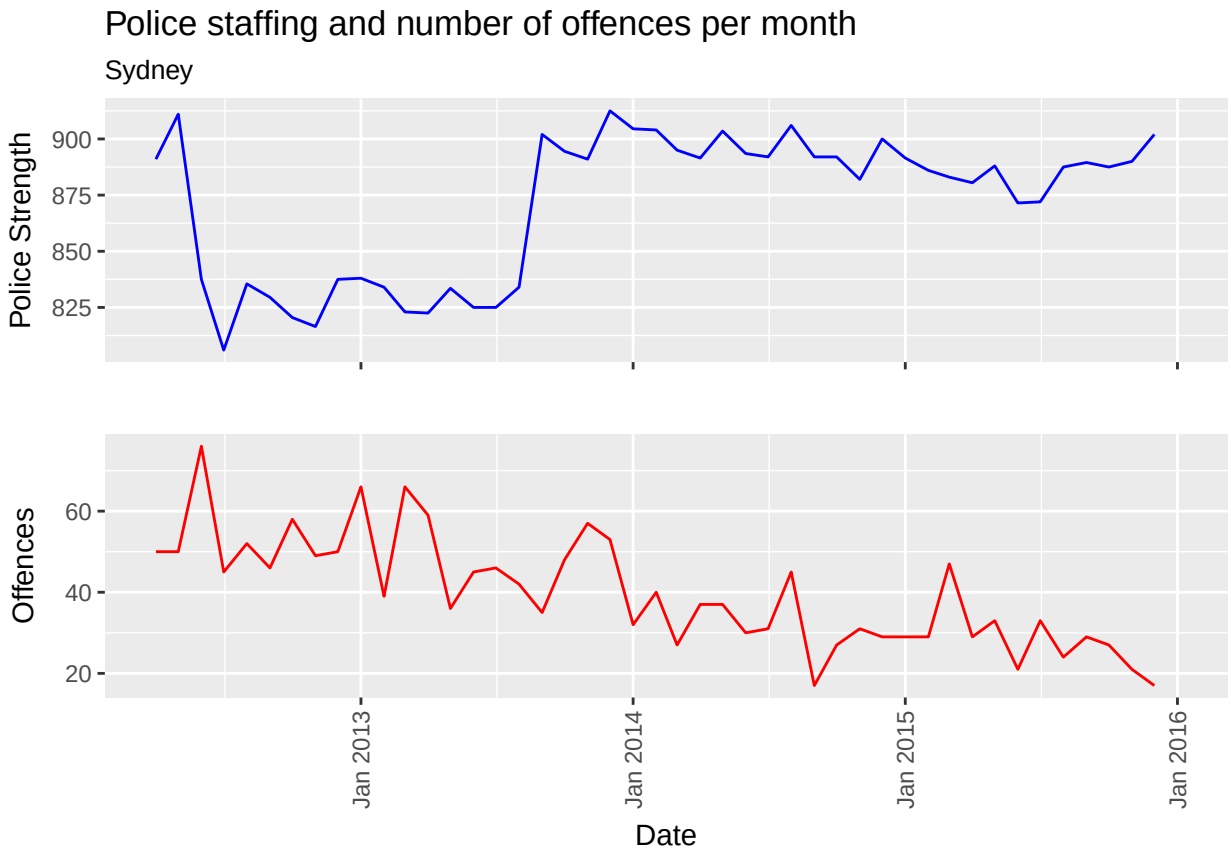

```
## [1] "LGA:  Sydney  Crime type:  Assault"
```



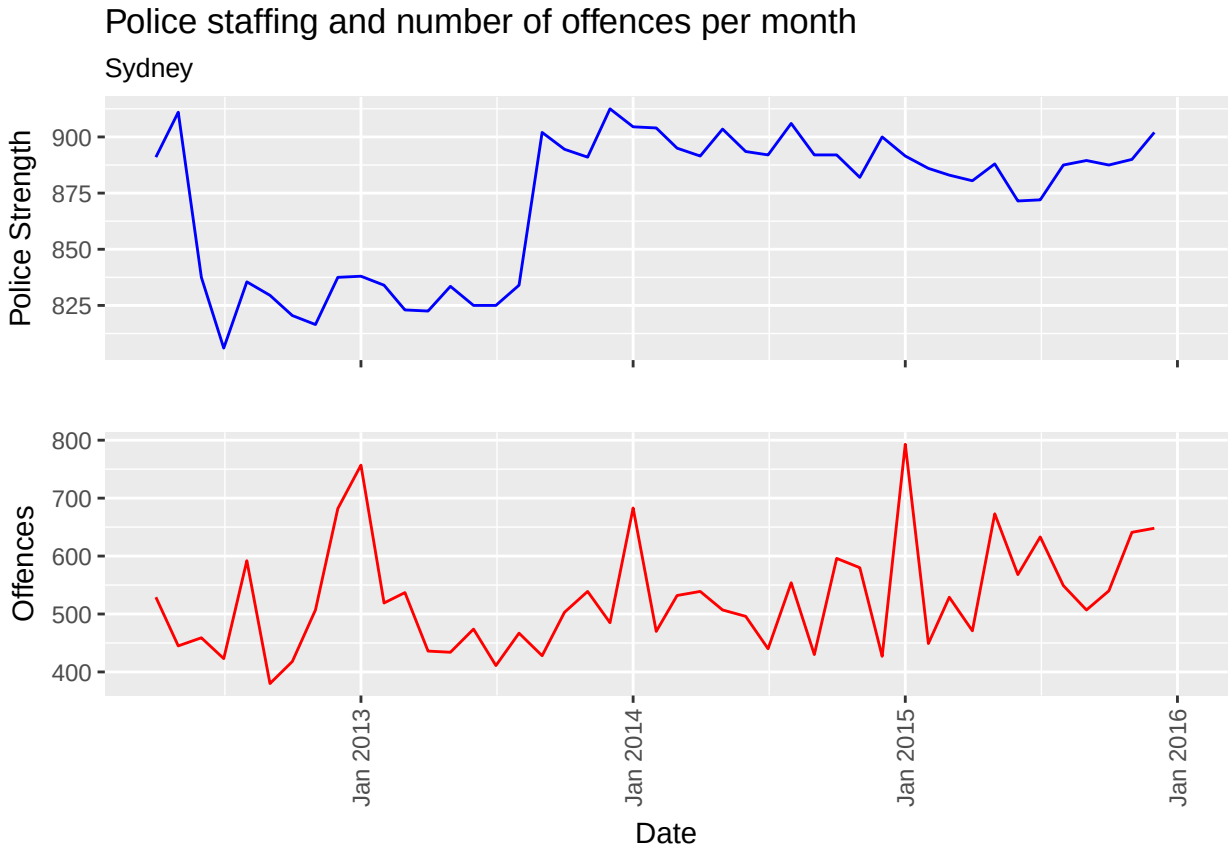

```
## [1] "LGA:  Sydney  Crime type:  Sexual Offences"
```



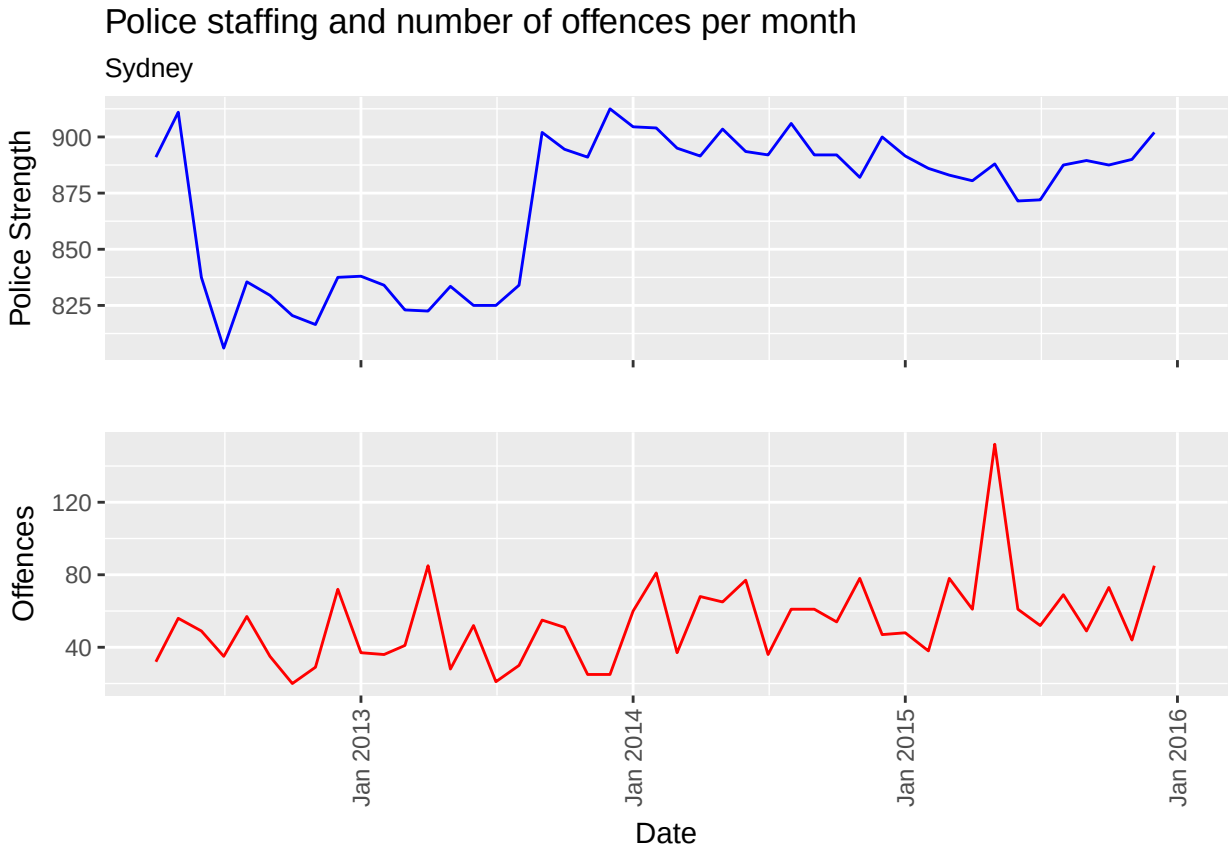

```
## [1] "LGA:  Sydney  Crime type:  Theft"
```



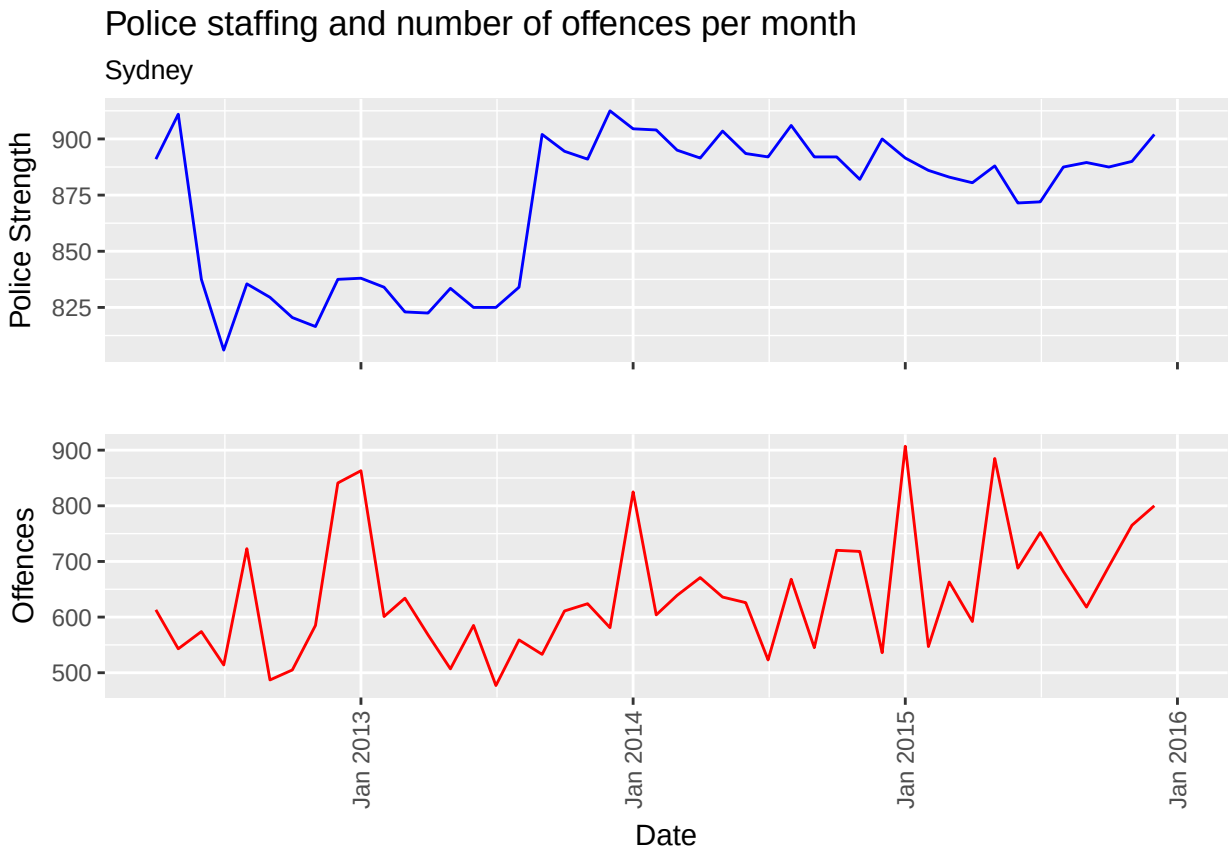

```
## [1] "LGA:  Sydney  Crime type:  Robbery"
```



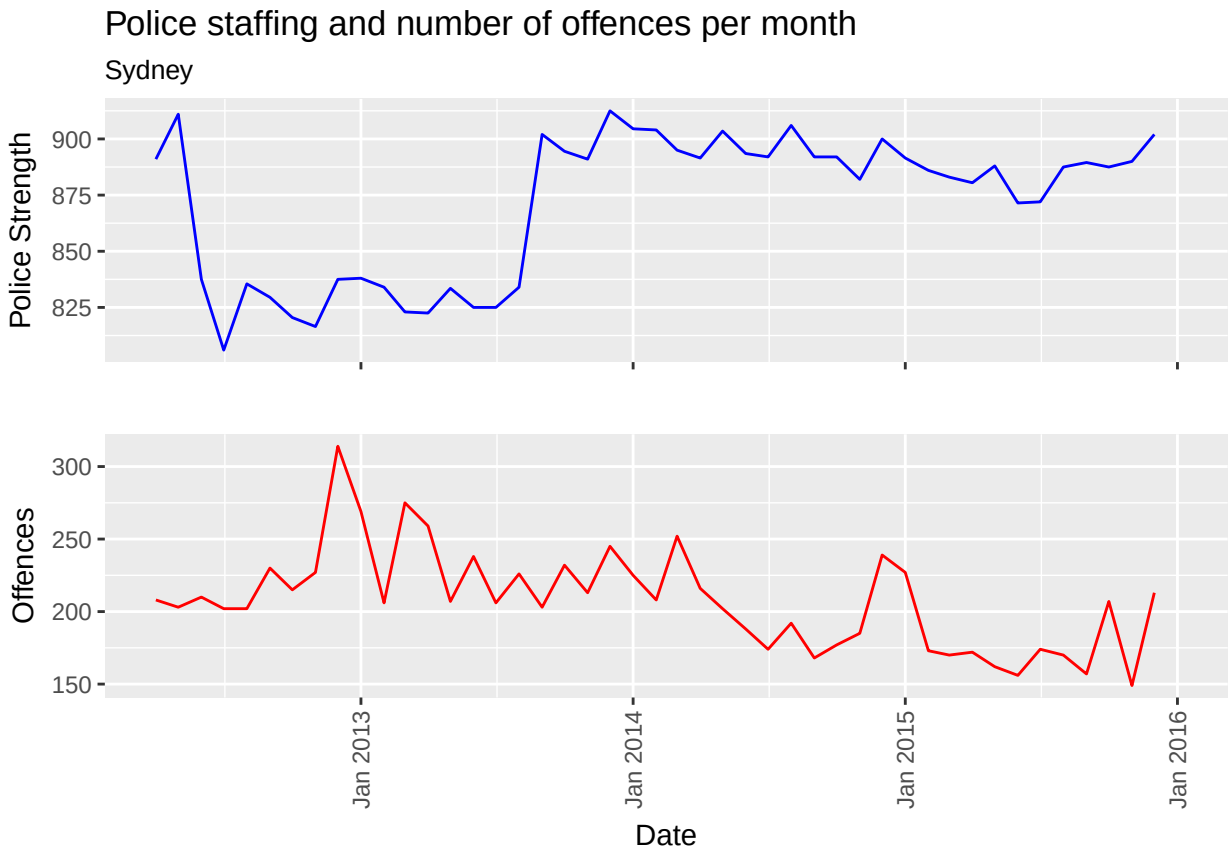

```
## [1] "LGA:  Sydney  Crime type:  Drug Possession"
```



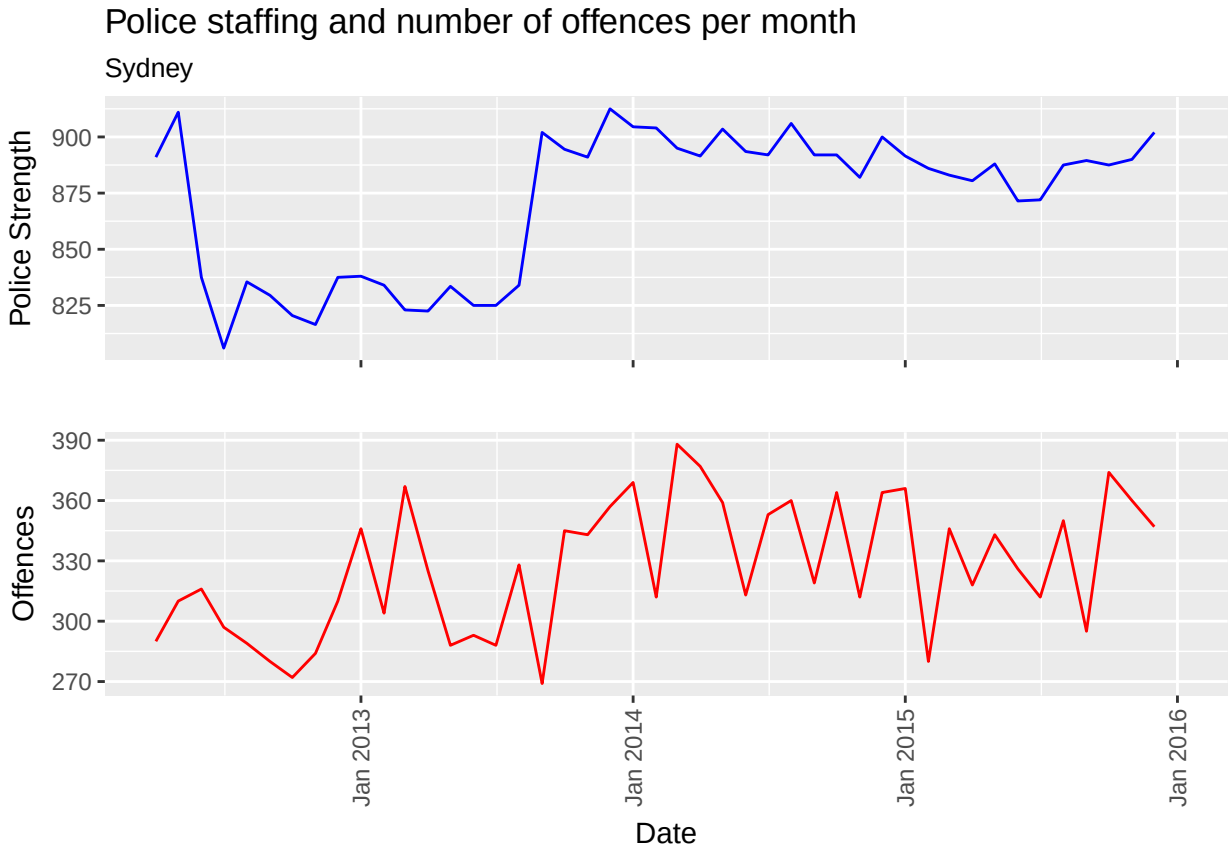

```
## [1] "LGA:  Sydney  Crime type:  Drug Dealing"
```



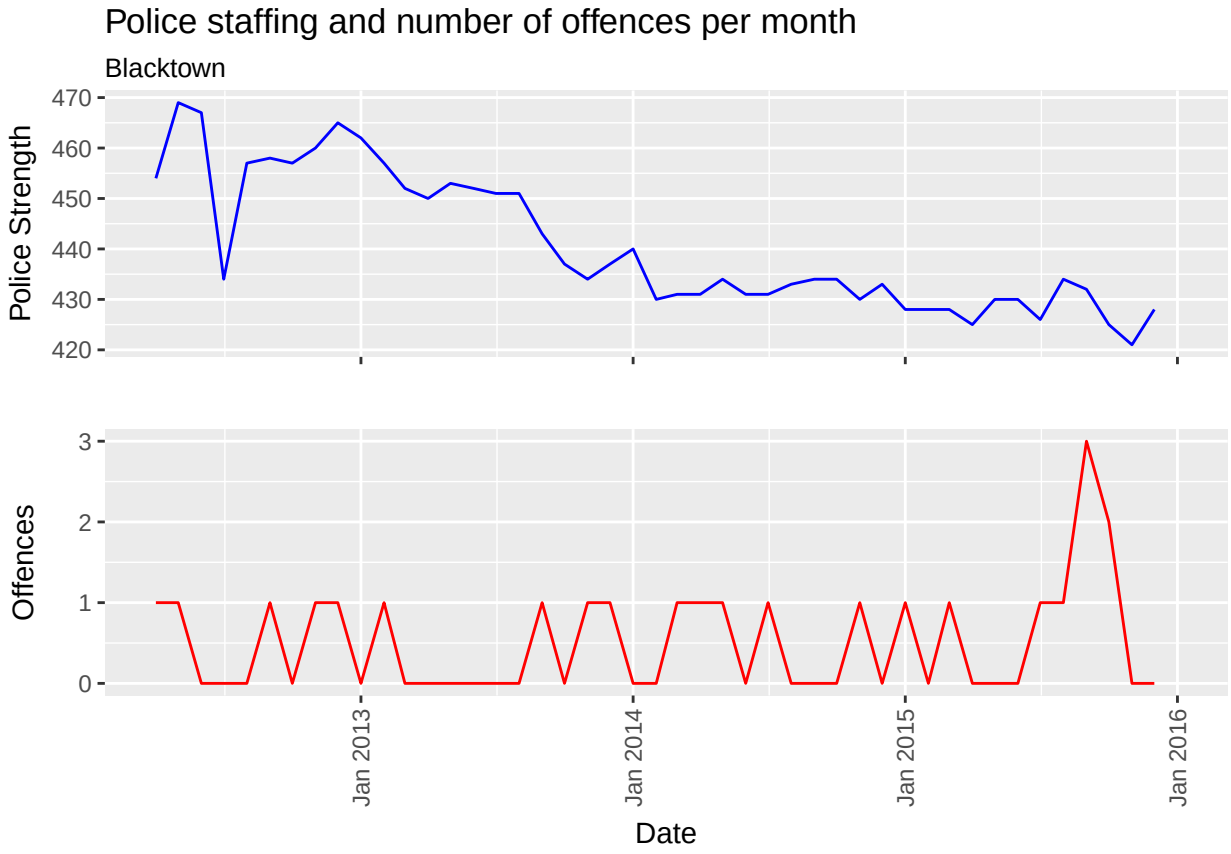

```
## [1] "LGA:  Sydney  Crime type:  Drug Offences"
```



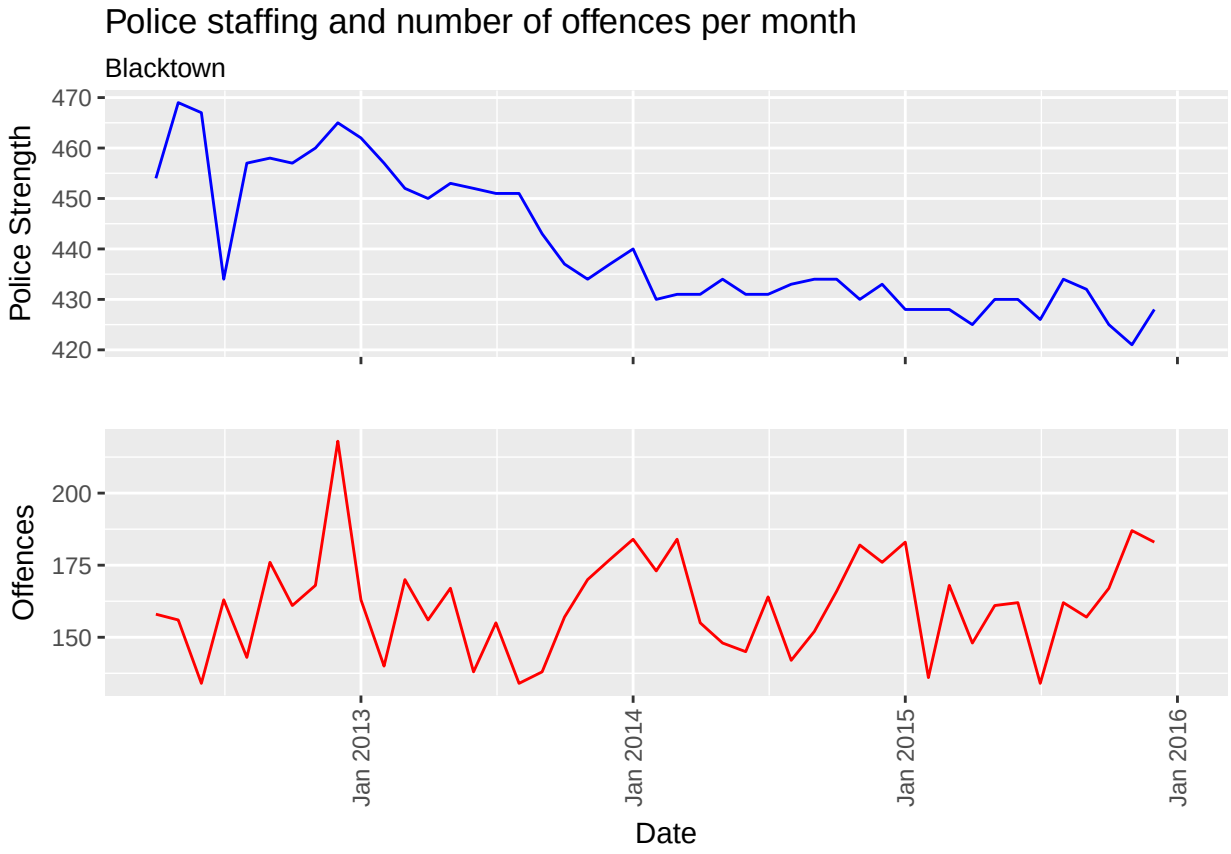

```
## [1] "LGA:  Sydney  Crime type:  Disorderly Conduct"
```



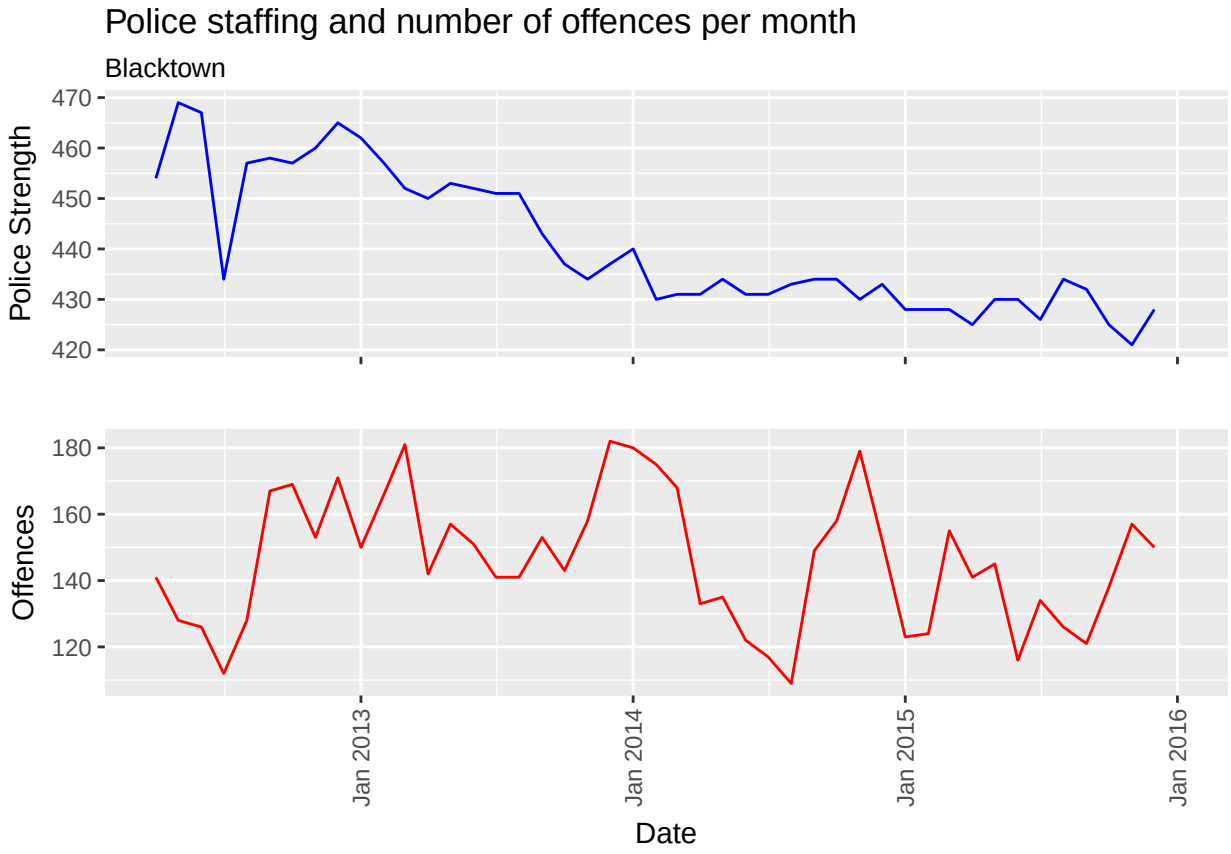

```
## [1] "LGA:  Sydney  Crime type:  Against Justice Procedures"
```



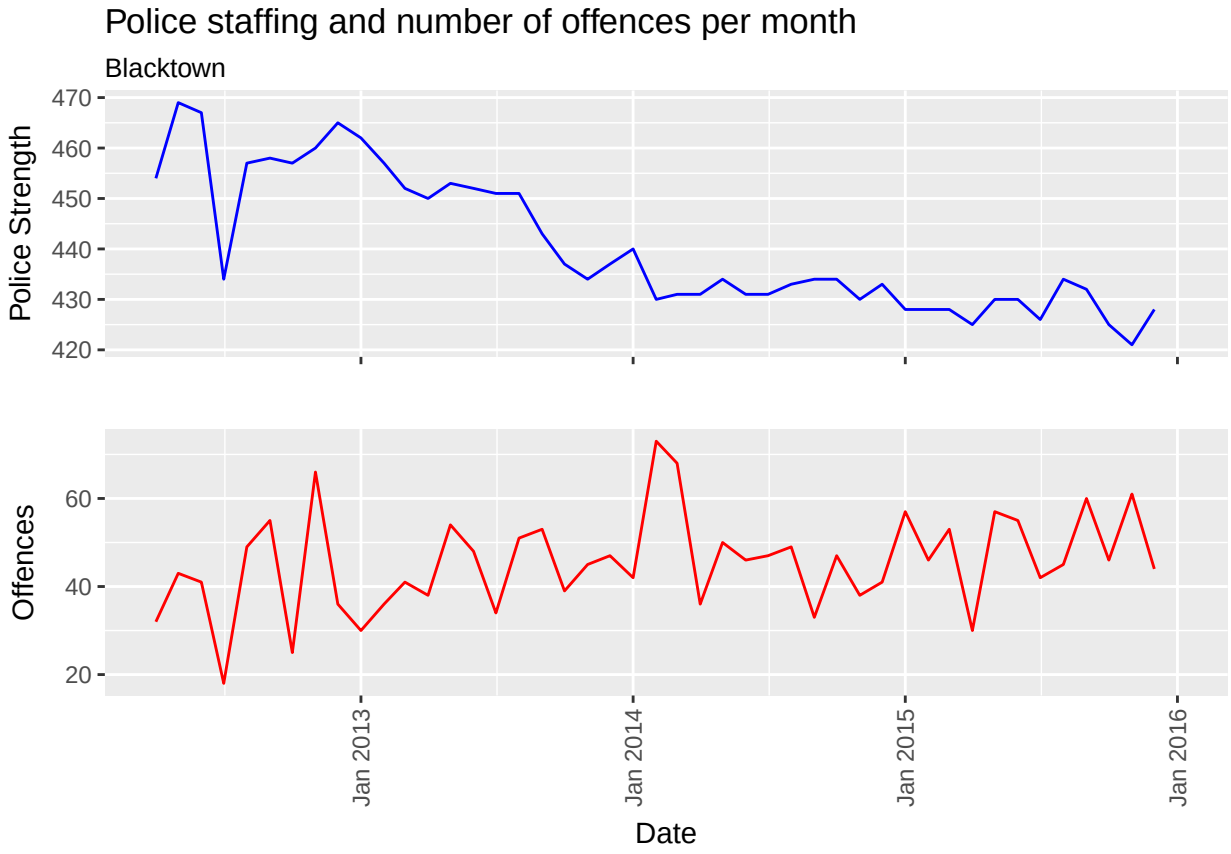

```
## [1] "LGA:  Blacktown  Crime type:  Homicide"
```



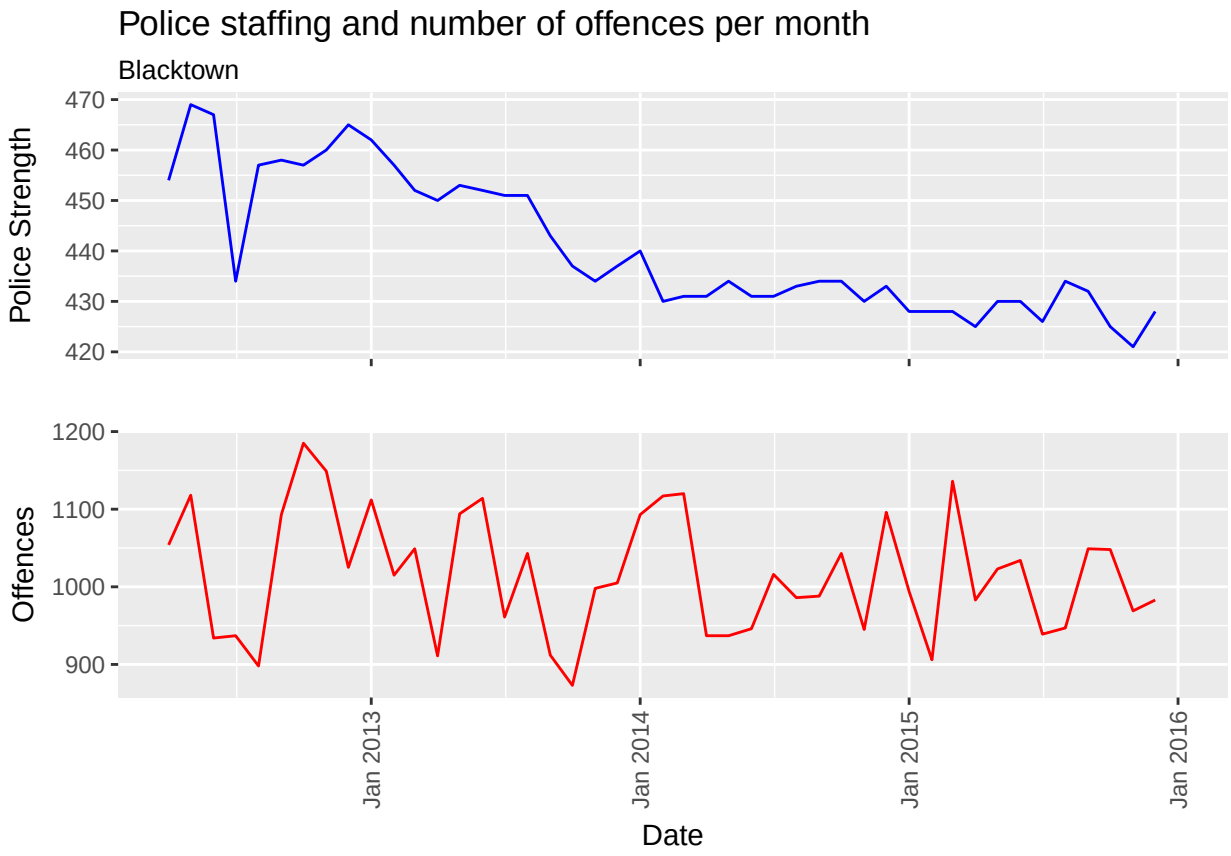

```
## [1] "LGA:  Blacktown  Crime type:  Domestic Assault"
```



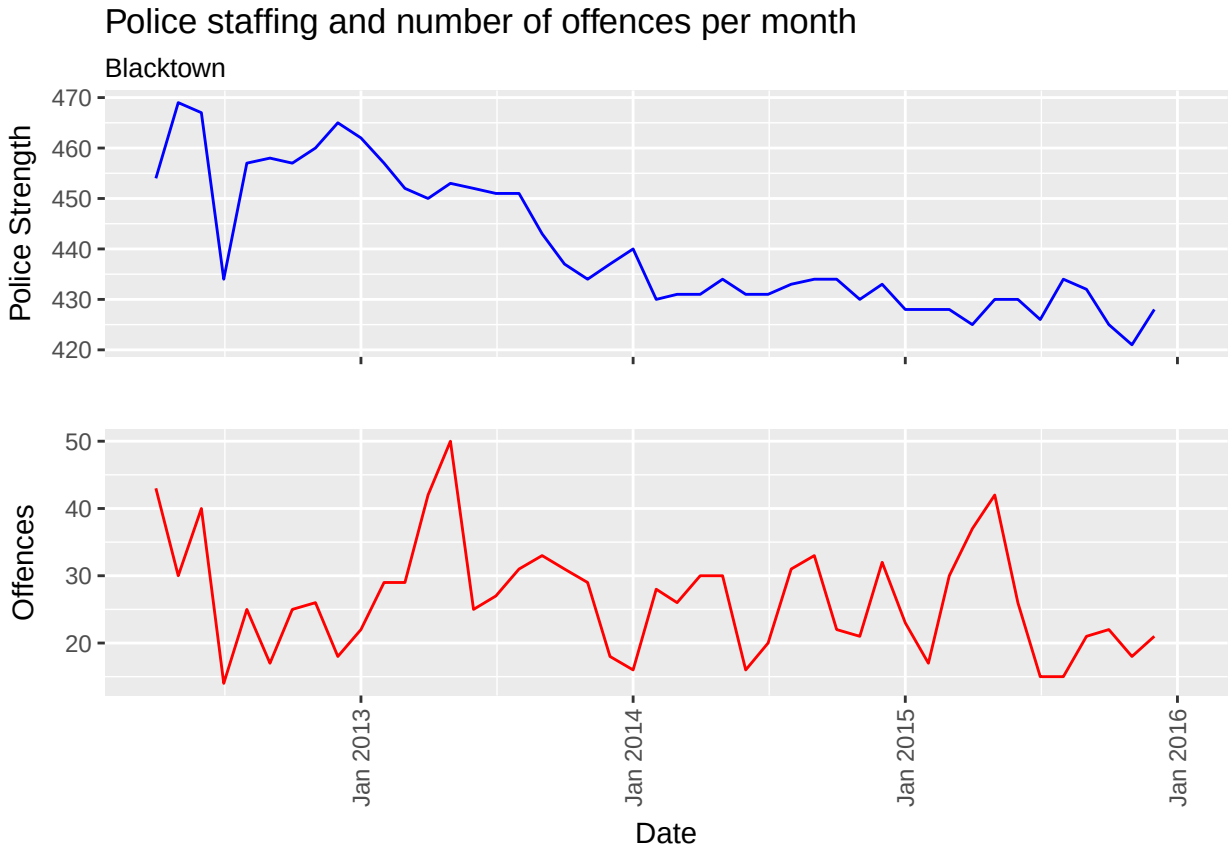

```
## [1] "LGA:  Blacktown  Crime type:  Assault"
```



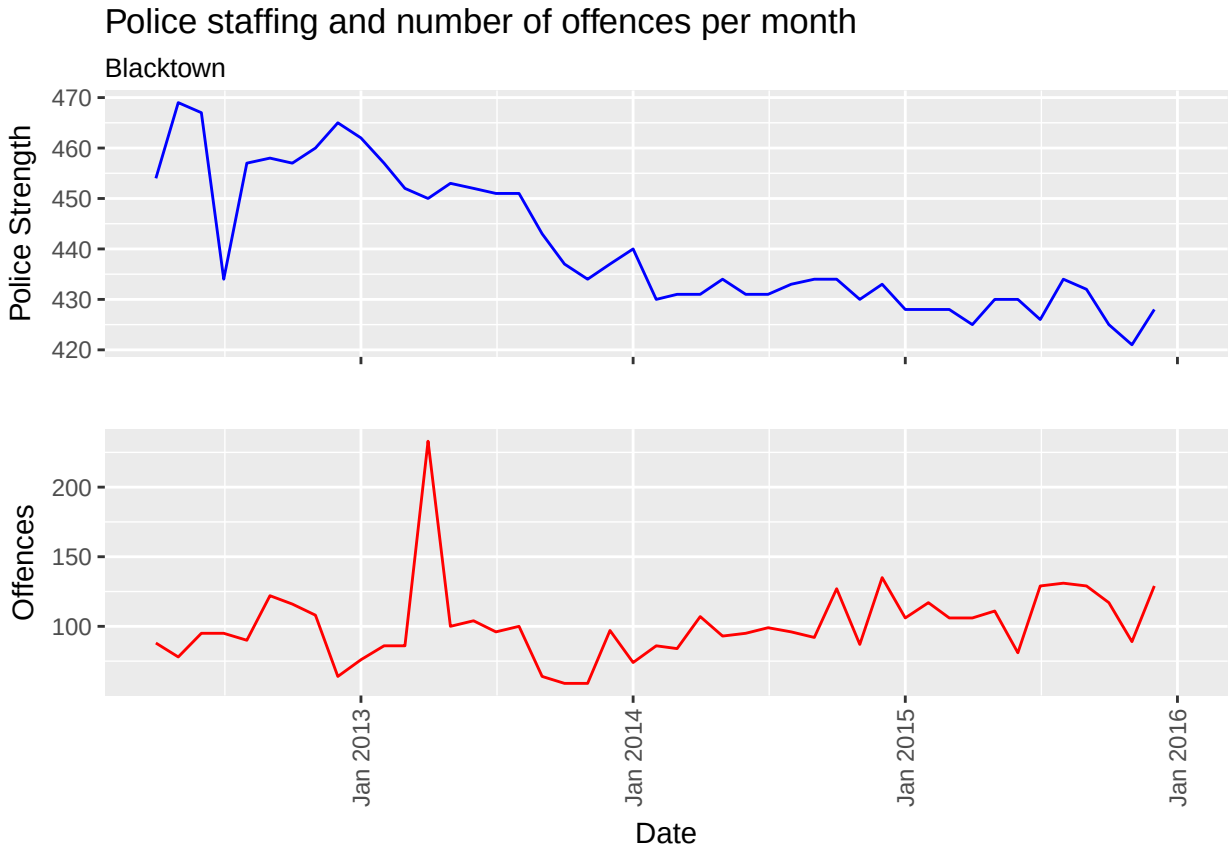

```
## [1] "LGA:  Blacktown  Crime type:  Sexual Offences"
```



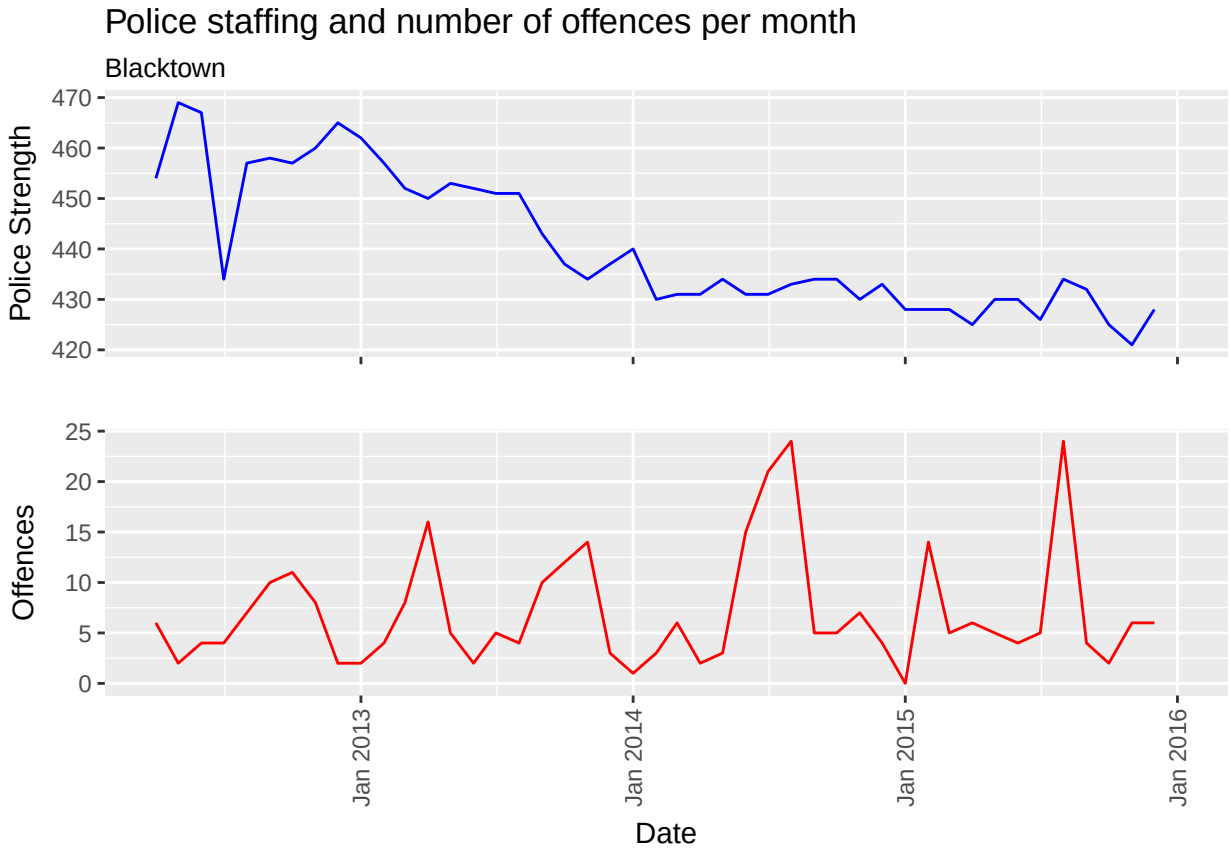

```
## [1] "LGA:  Blacktown  Crime type:  Theft"
```



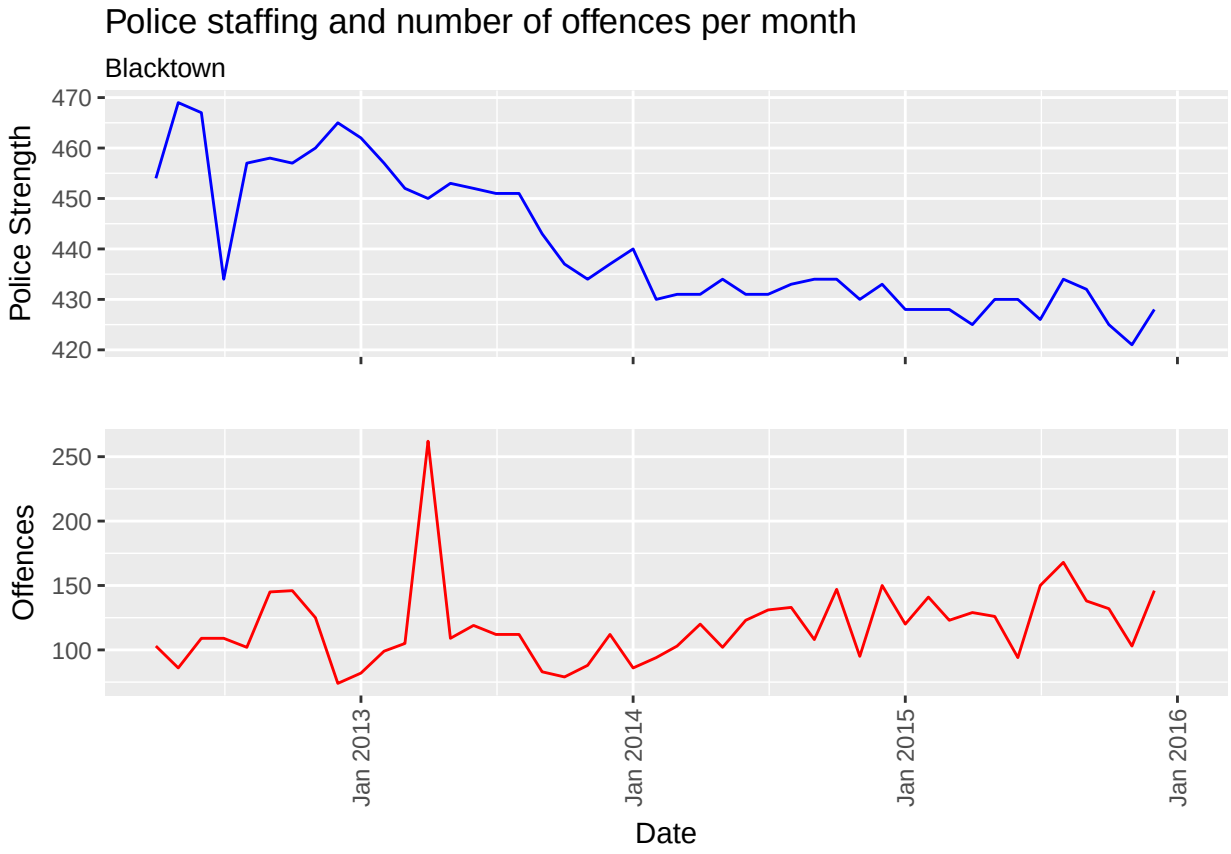

```
## [1] "LGA:  Blacktown  Crime type:  Robbery"
```



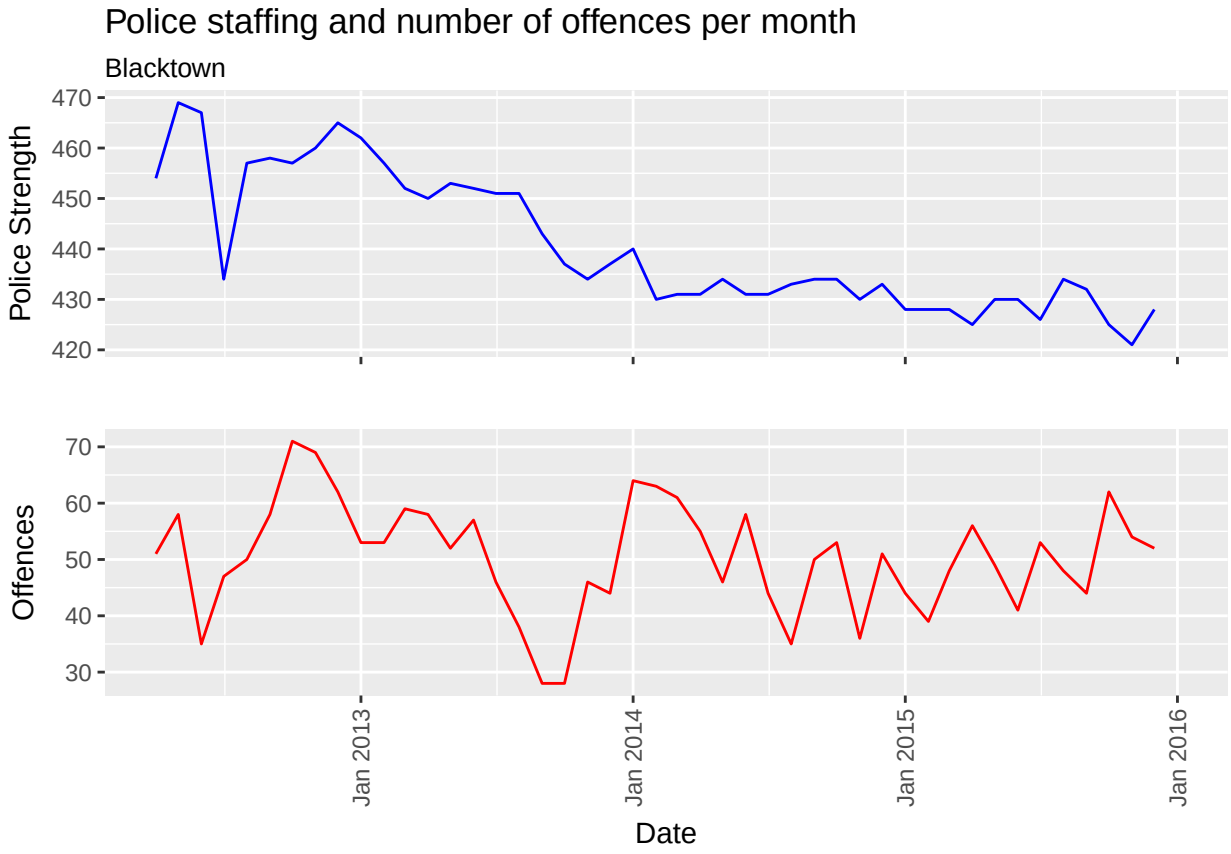

```
## [1] "LGA:  Blacktown  Crime type:  Drug Possession"
```



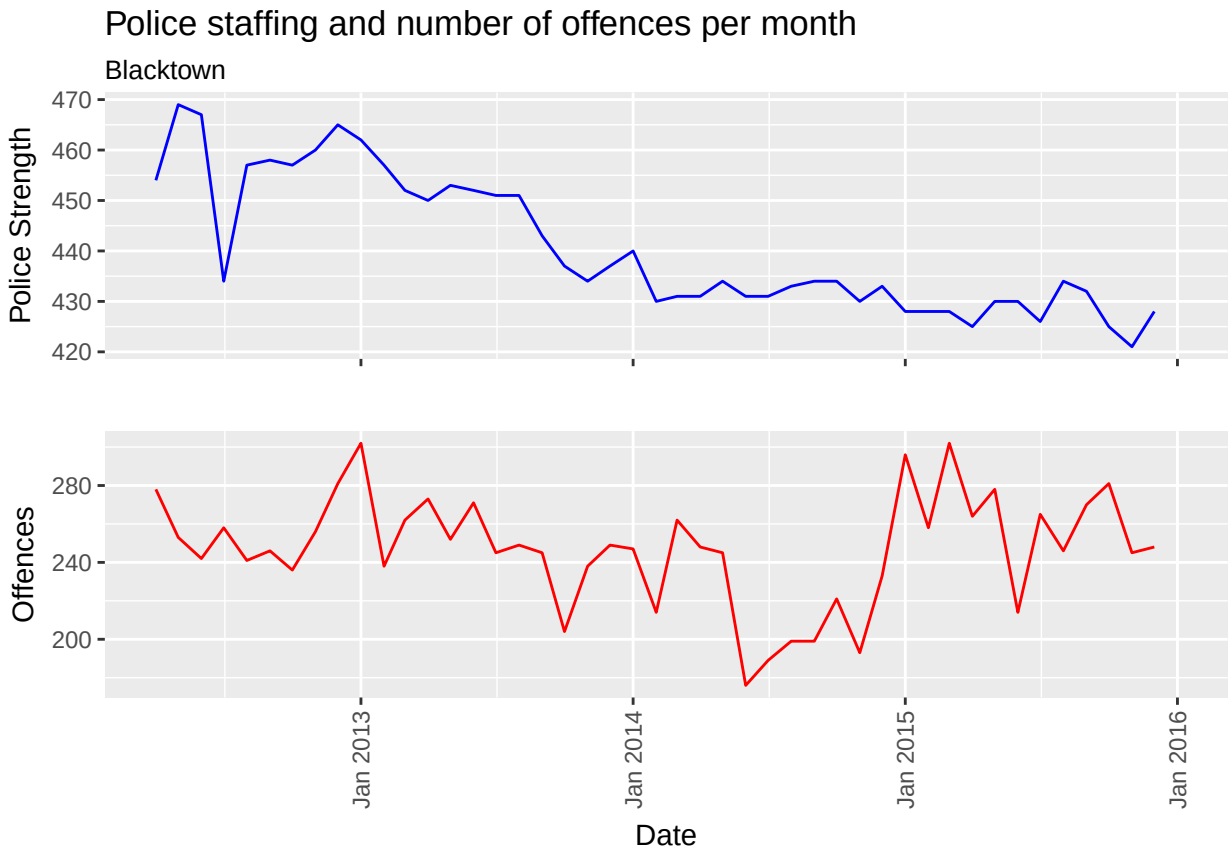

```
## [1] "LGA:  Blacktown  Crime type:  Drug Dealing"
```



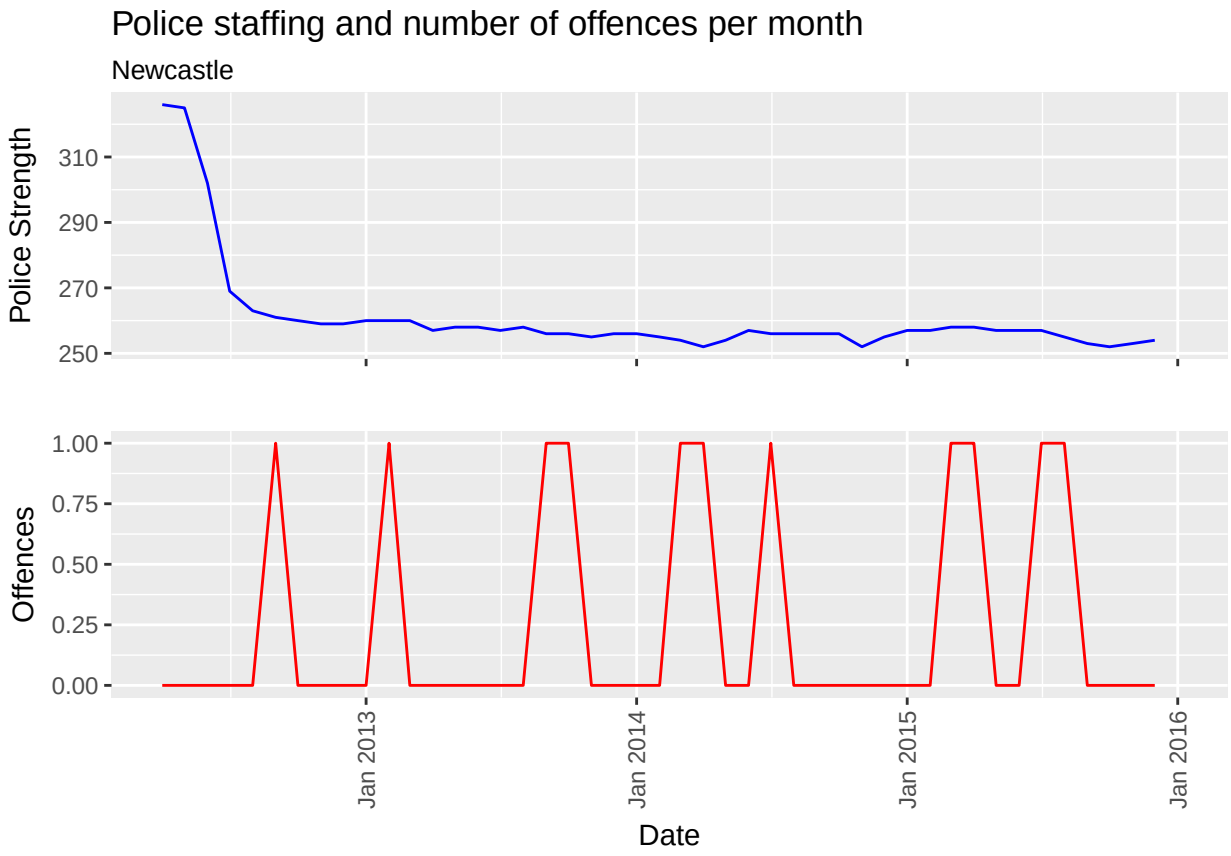

```
## [1] "LGA:  Blacktown  Crime type:  Drug Offences"
```



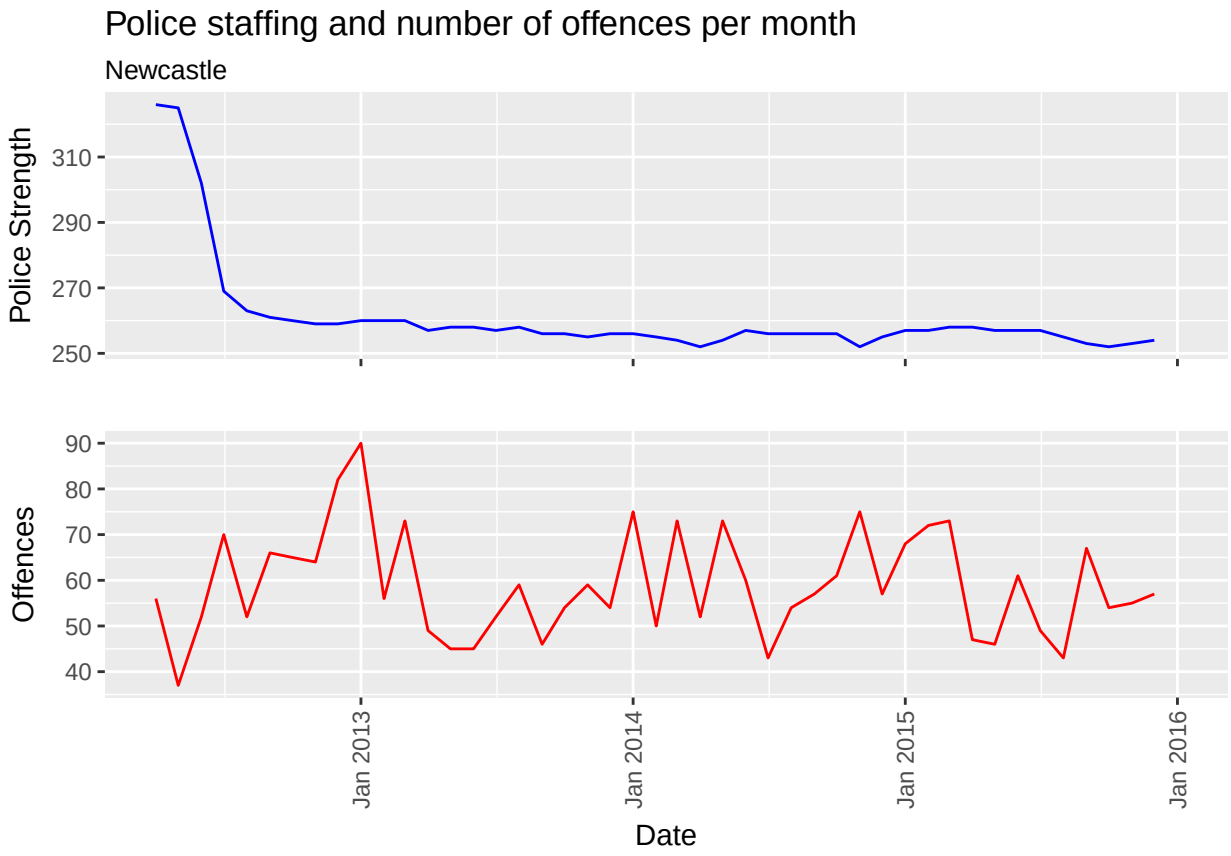

```
## [1] "LGA:  Blacktown  Crime type:  Disorderly Conduct"
```



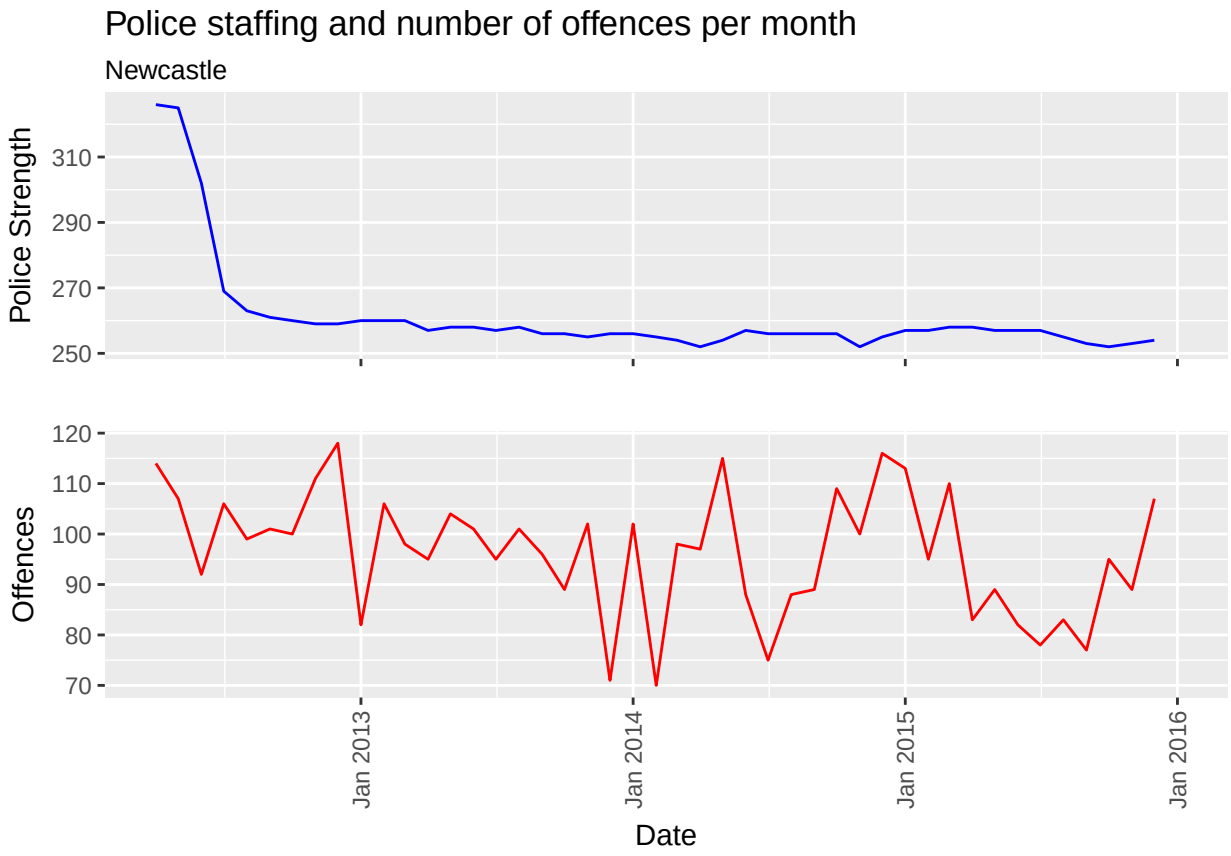

```
## [1] "LGA:  Blacktown  Crime type:  Against Justice Procedures"
```



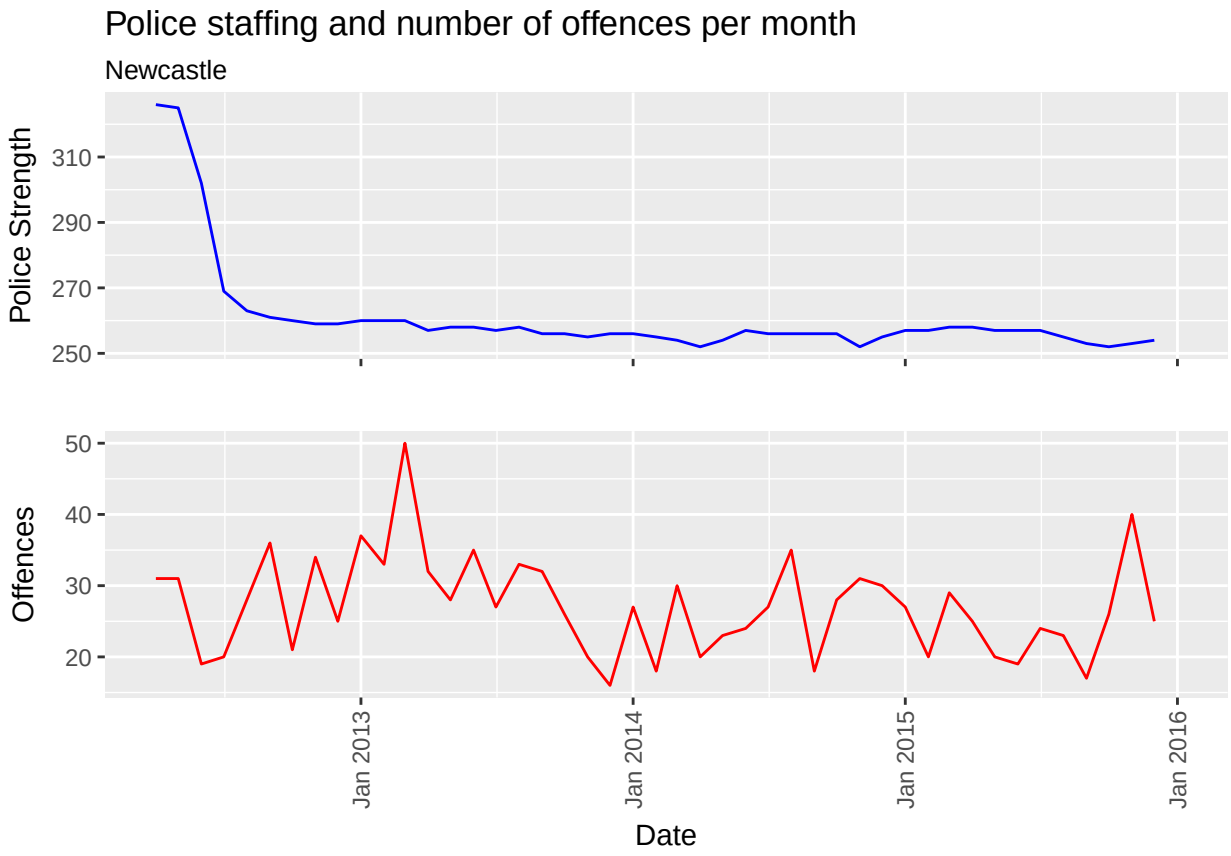

```
## [1] "LGA:  Newcastle  Crime type:  Homicide"
```



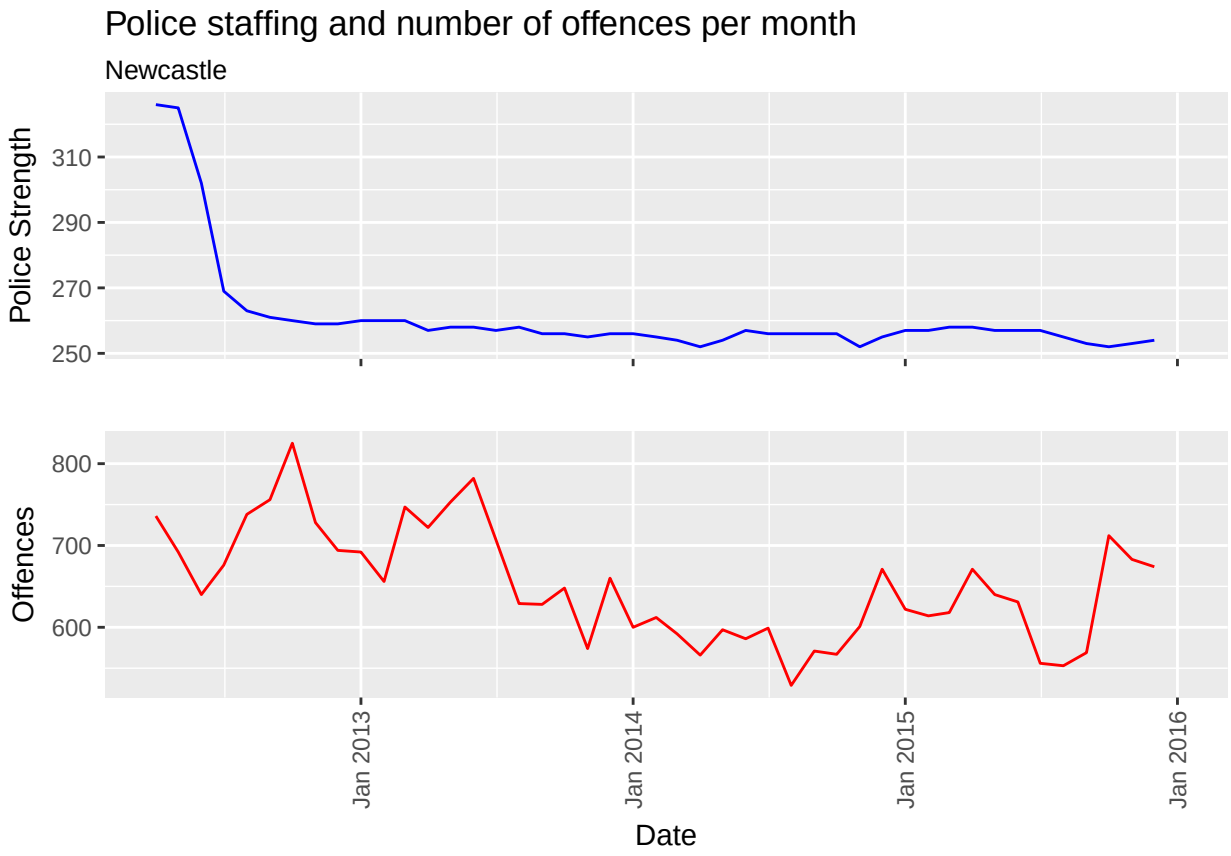

```
## [1] "LGA:  Newcastle  Crime type:  Domestic Assault"
```



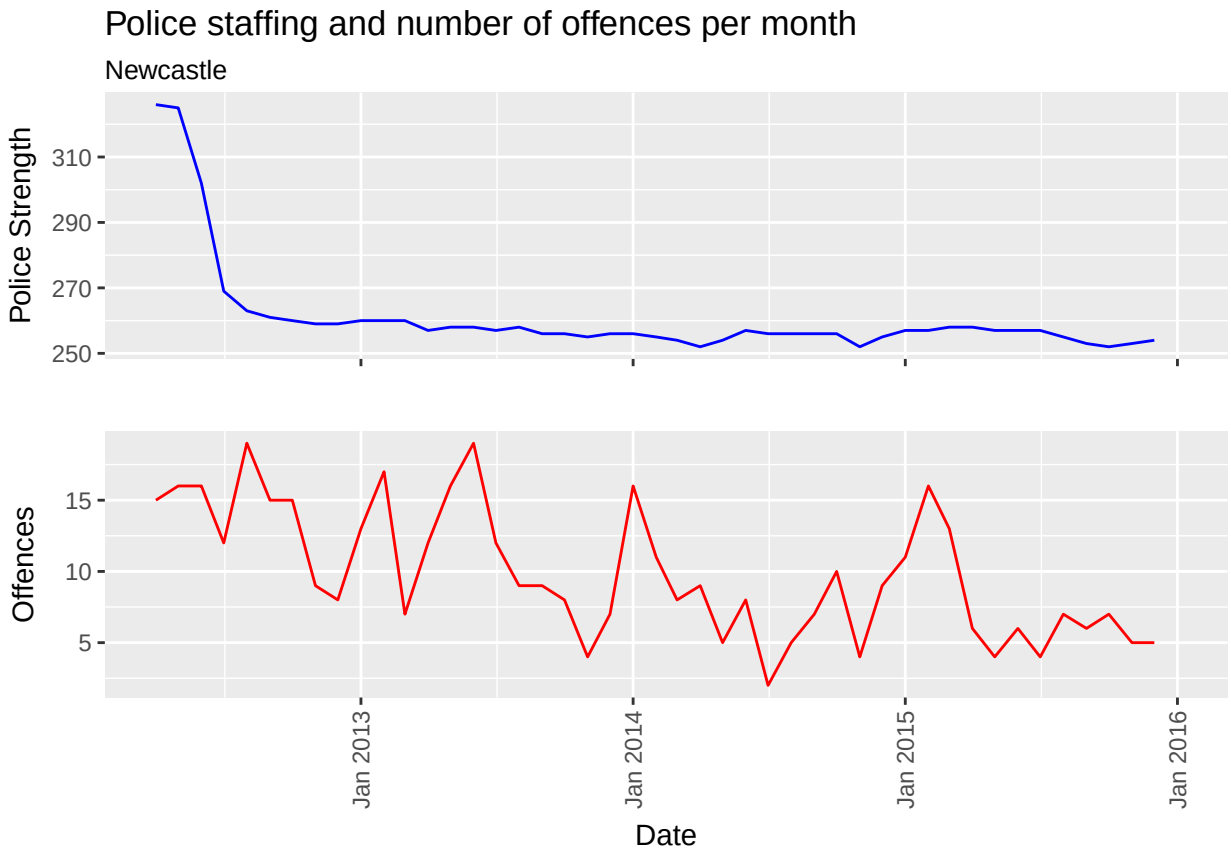

```
## [1] "LGA:  Newcastle  Crime type:  Assault"
```



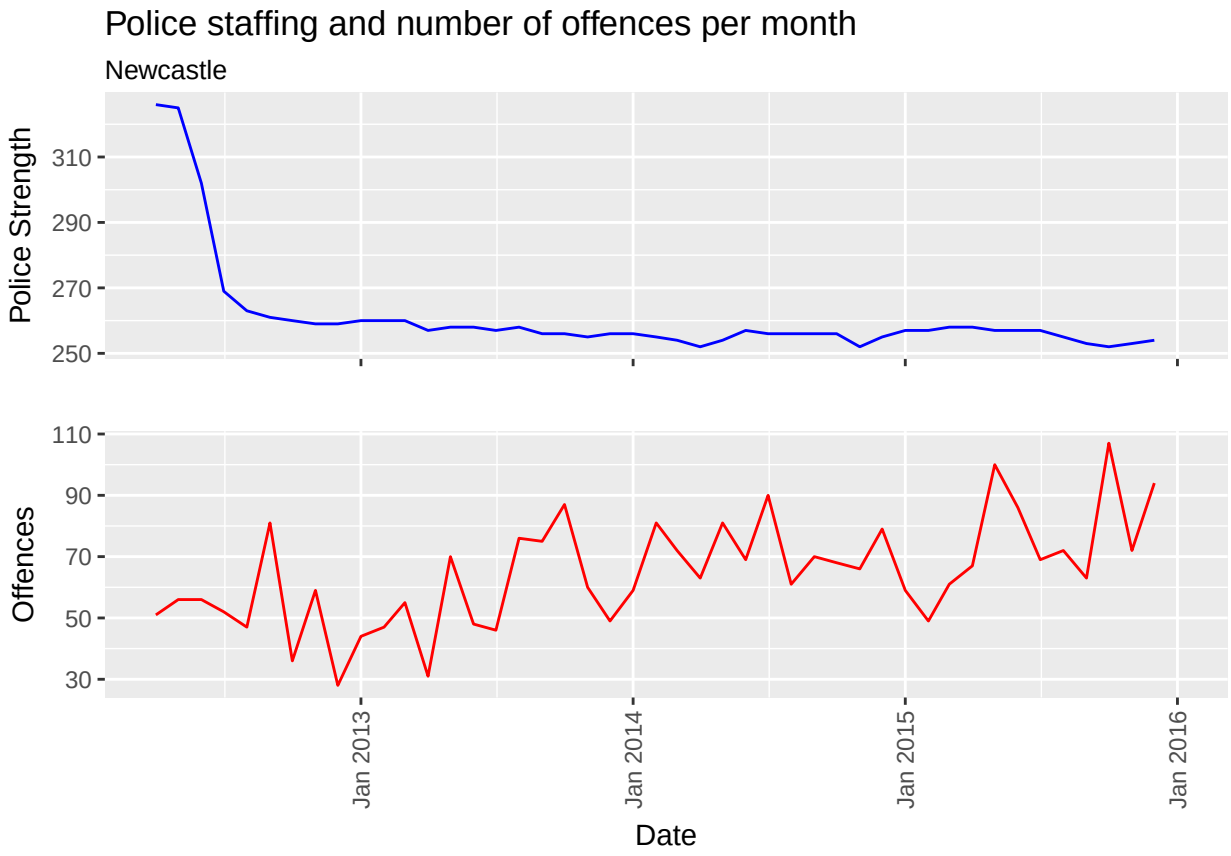

```
## [1] "LGA:  Newcastle  Crime type:  Sexual Offences"
```



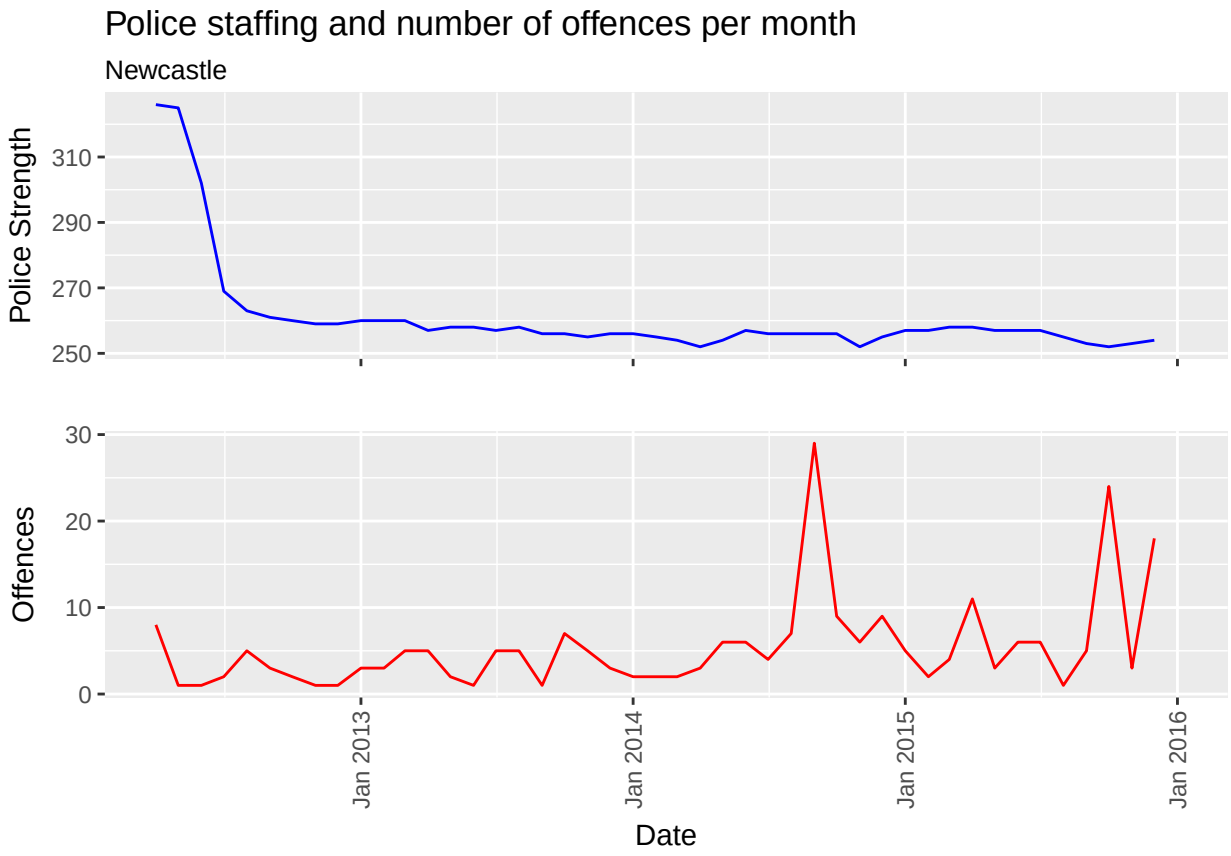

```
## [1] "LGA:  Newcastle  Crime type:  Theft"
```



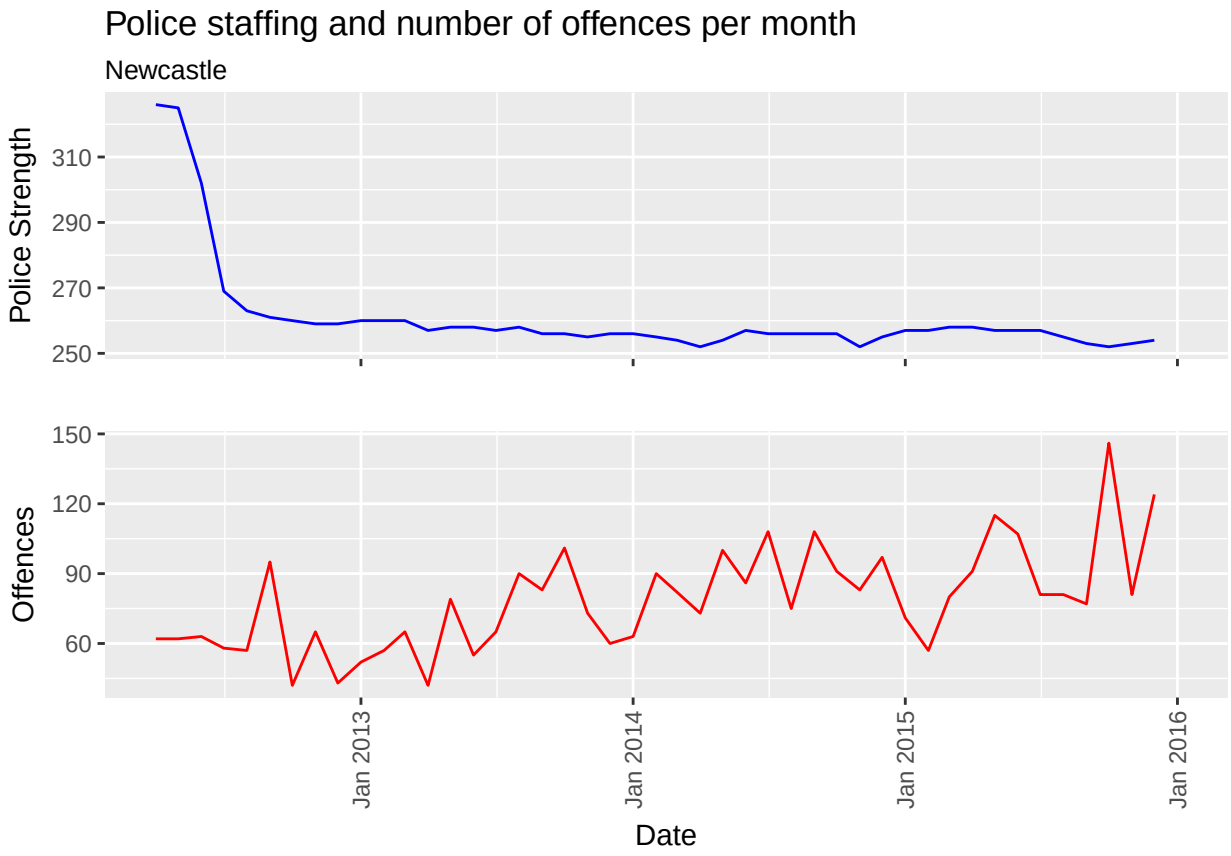

```
## [1] "LGA:  Newcastle  Crime type:  Robbery"
```



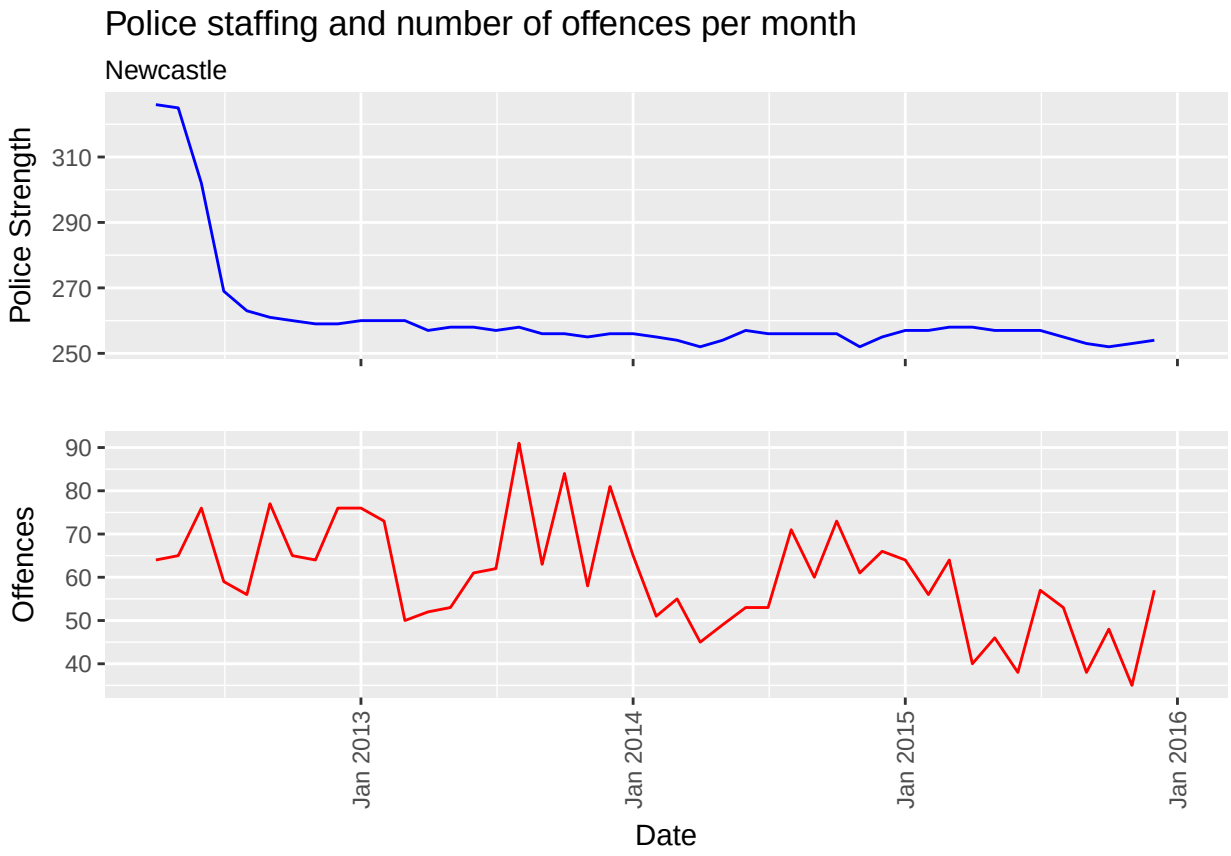

```
## [1] "LGA:  Newcastle  Crime type:  Drug Possession"
```



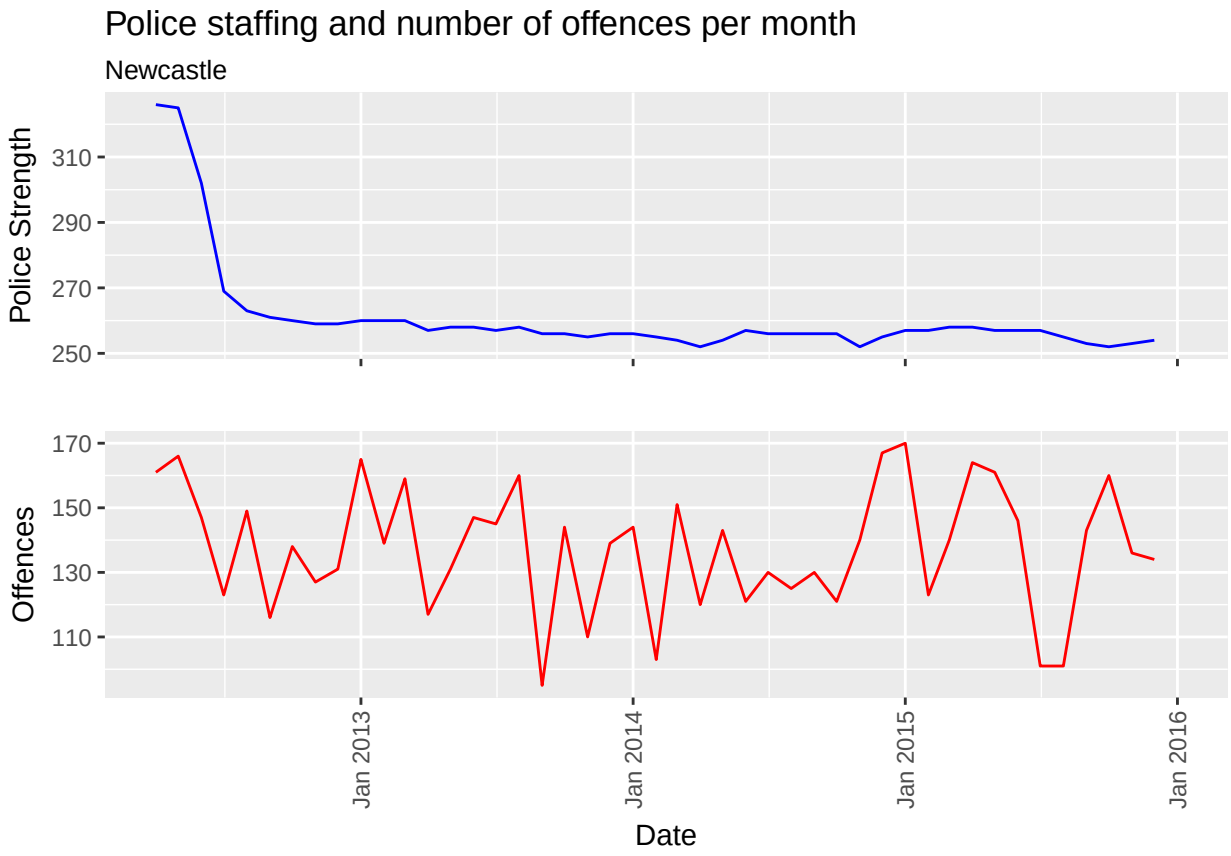

```
## [1] "LGA:  Newcastle  Crime type:  Drug Dealing"
```



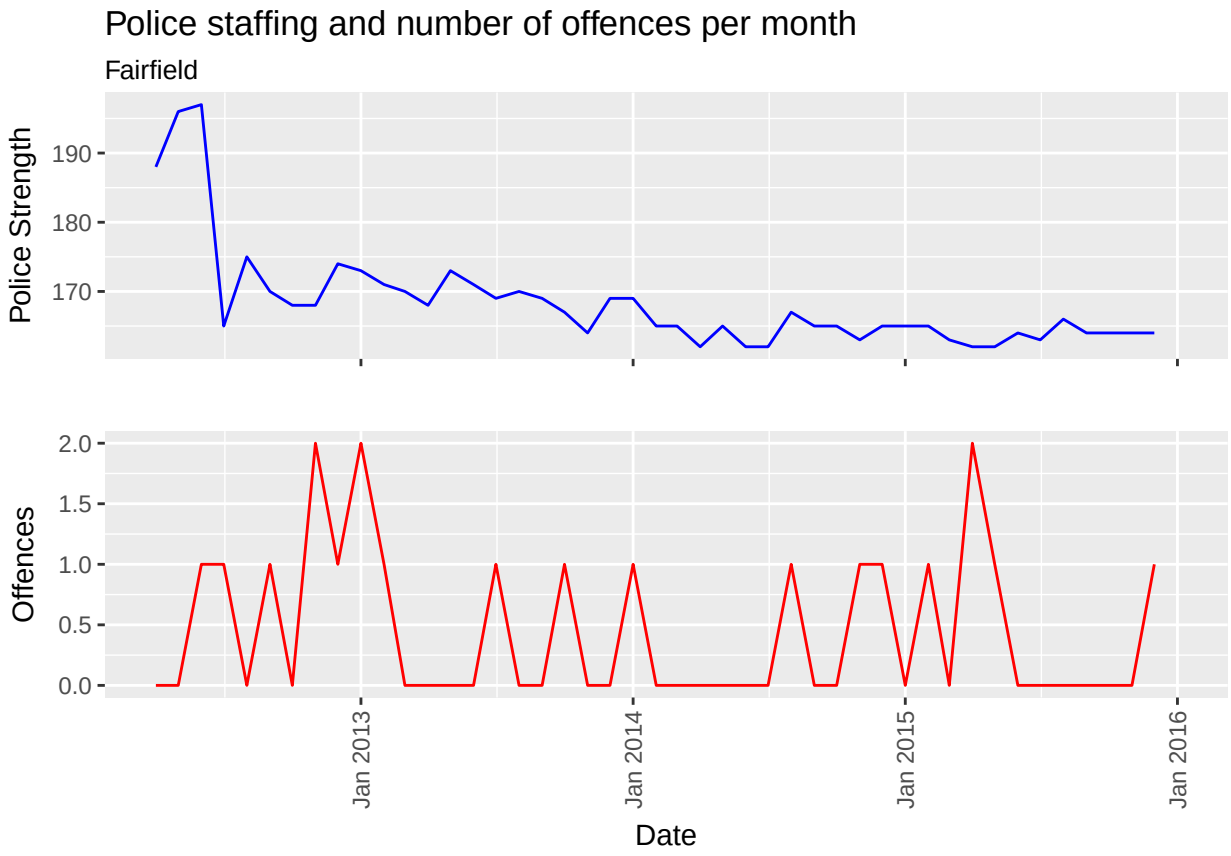

```
## [1] "LGA:  Newcastle  Crime type:  Drug Offences"
```



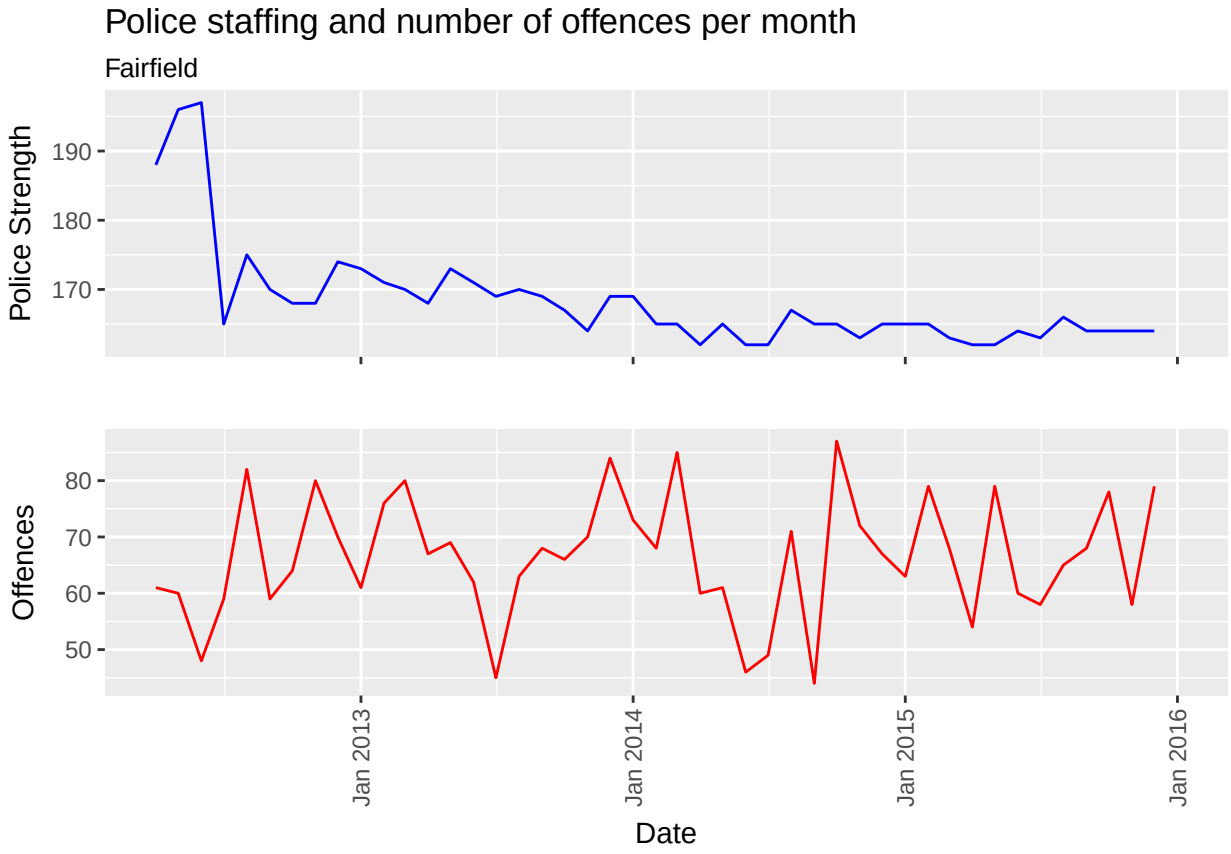

```
## [1] "LGA:  Newcastle  Crime type:  Disorderly Conduct"
```



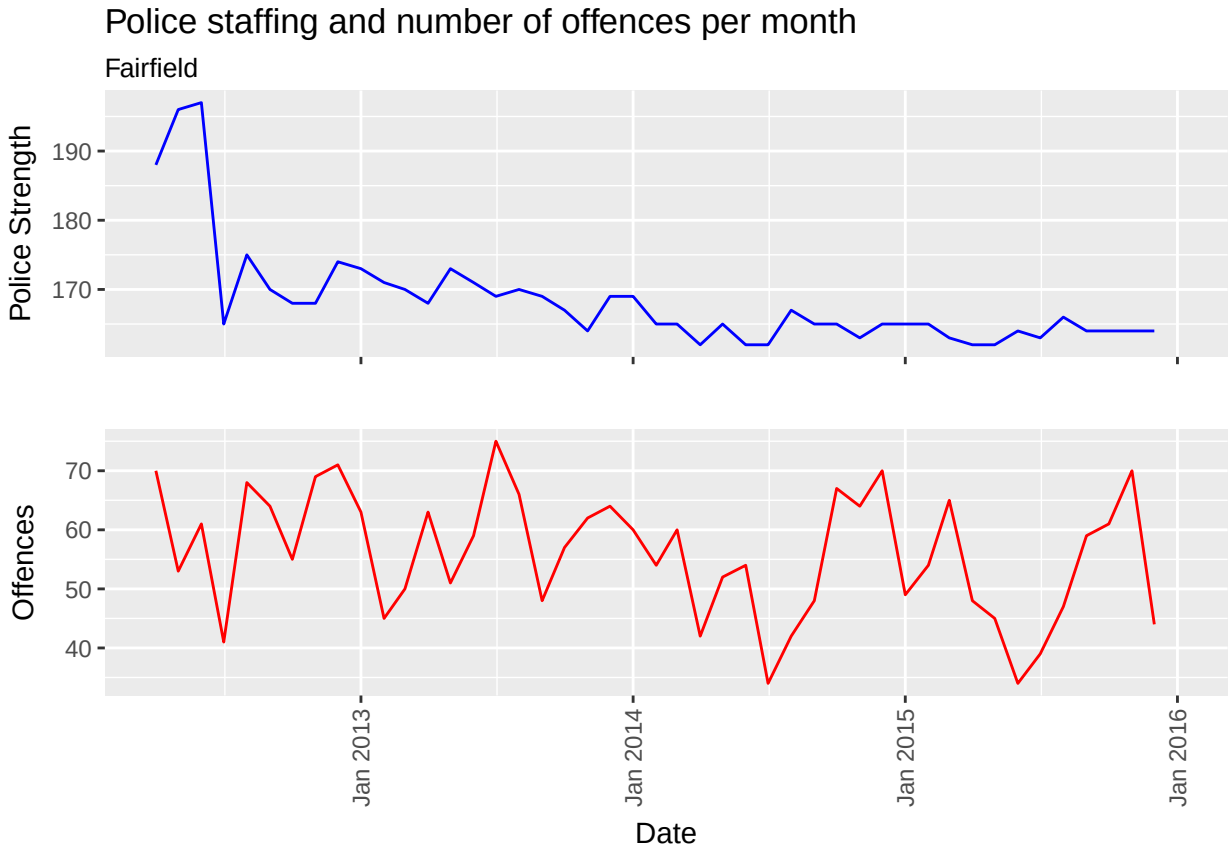

```
## [1] "LGA:  Newcastle  Crime type:  Against Justice Procedures"
```



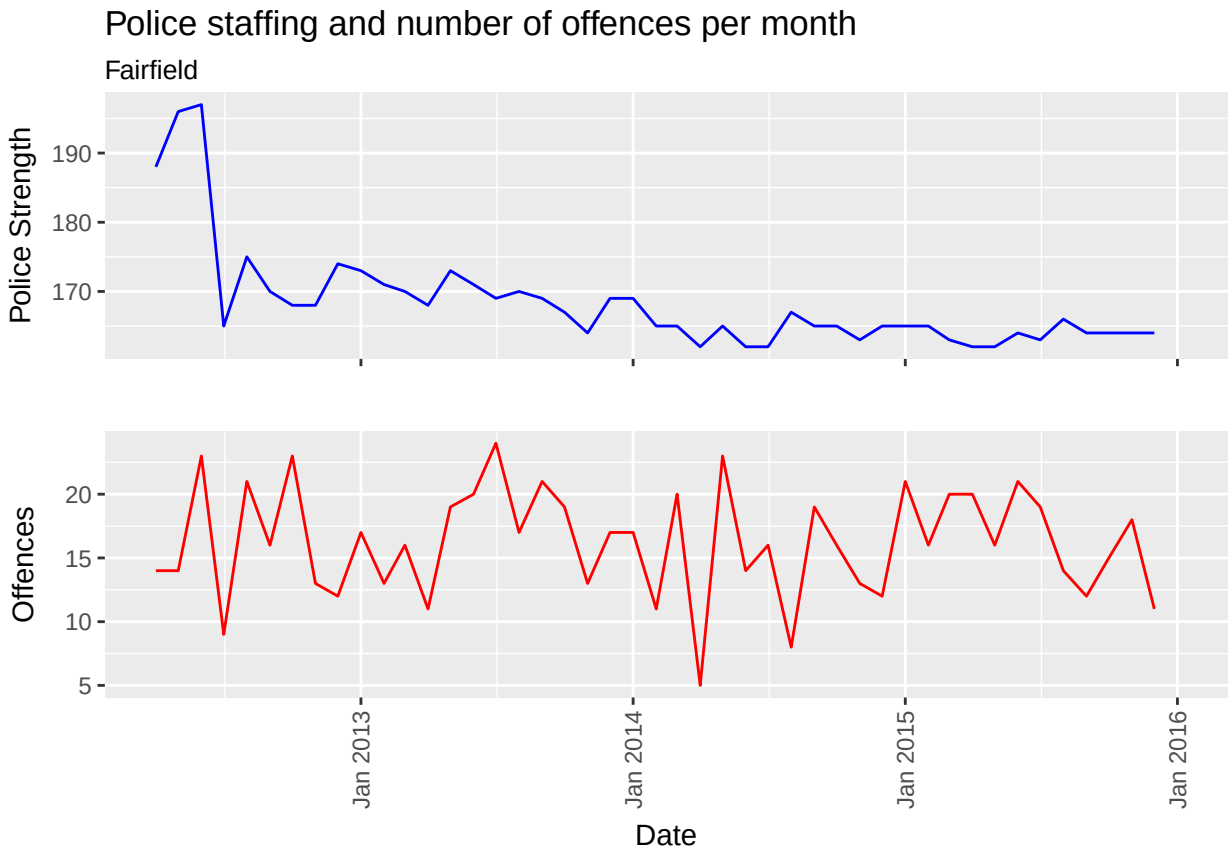

```
## [1] "LGA:  Fairfield  Crime type:  Homicide"
```



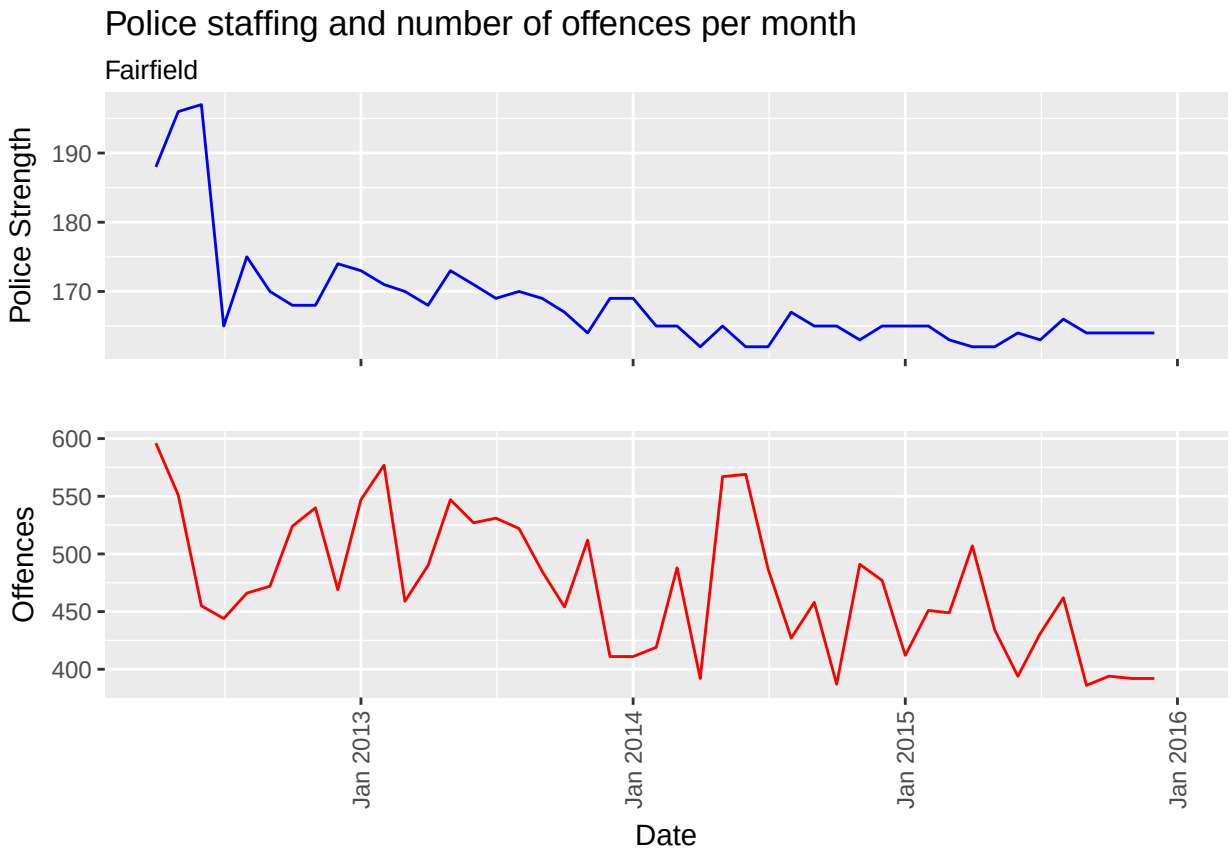

```
## [1] "LGA:  Fairfield  Crime type:  Domestic Assault"
```



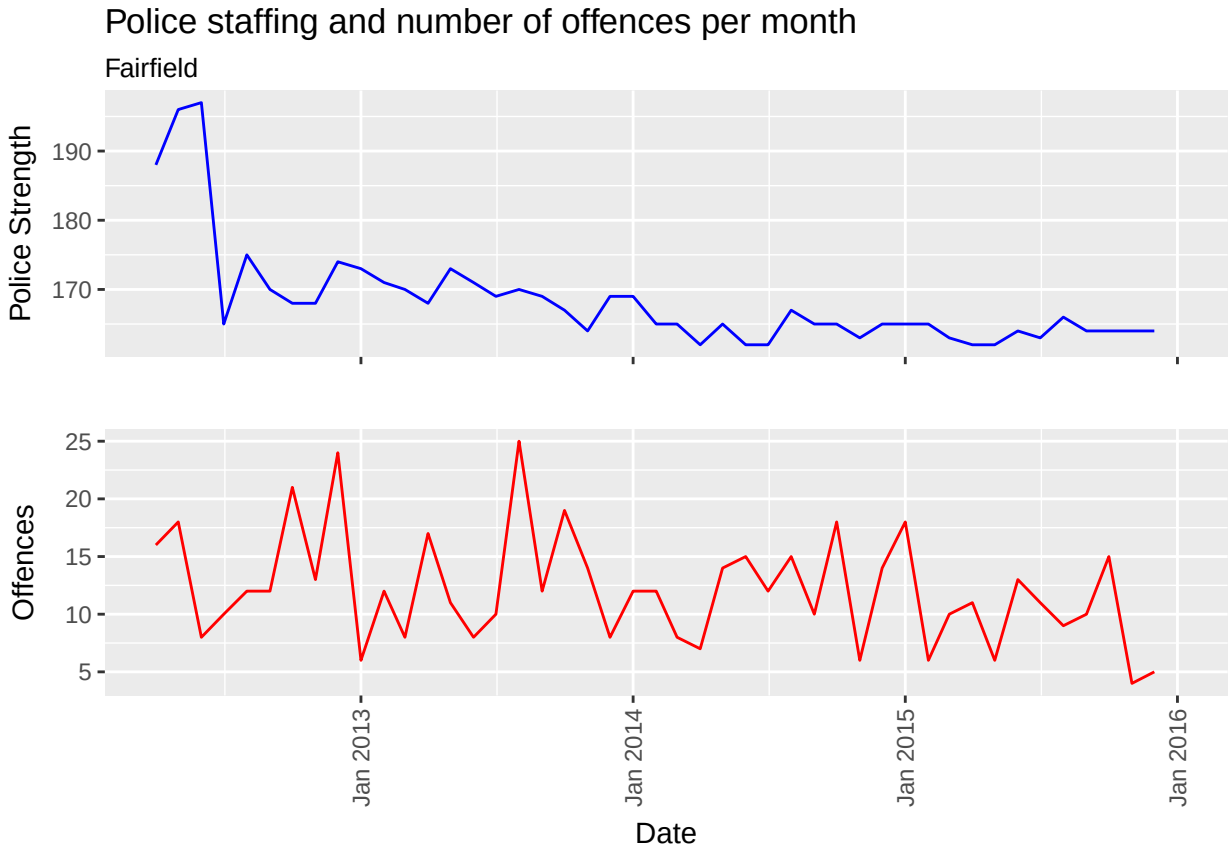

```
## [1] "LGA:  Fairfield  Crime type:  Assault"
```



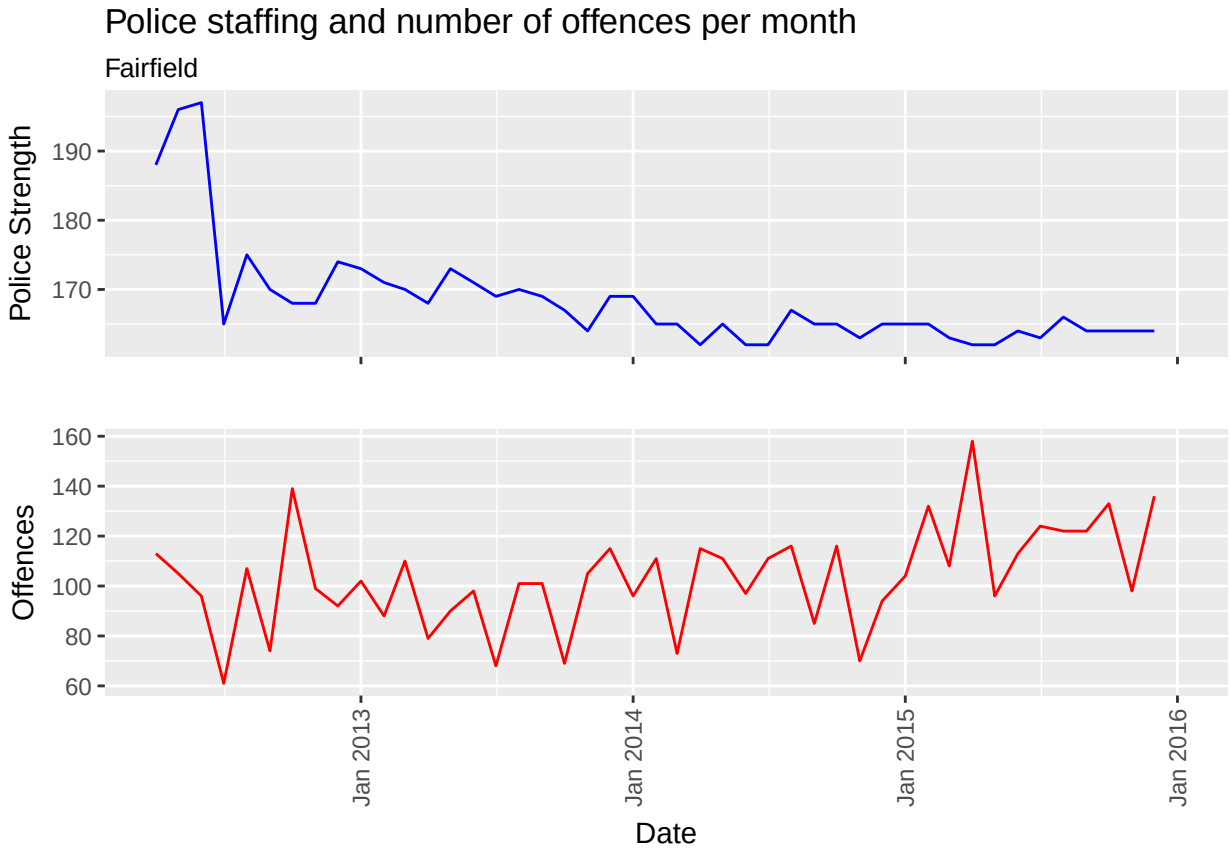

```
## [1] "LGA:  Fairfield  Crime type:  Sexual Offences"
```



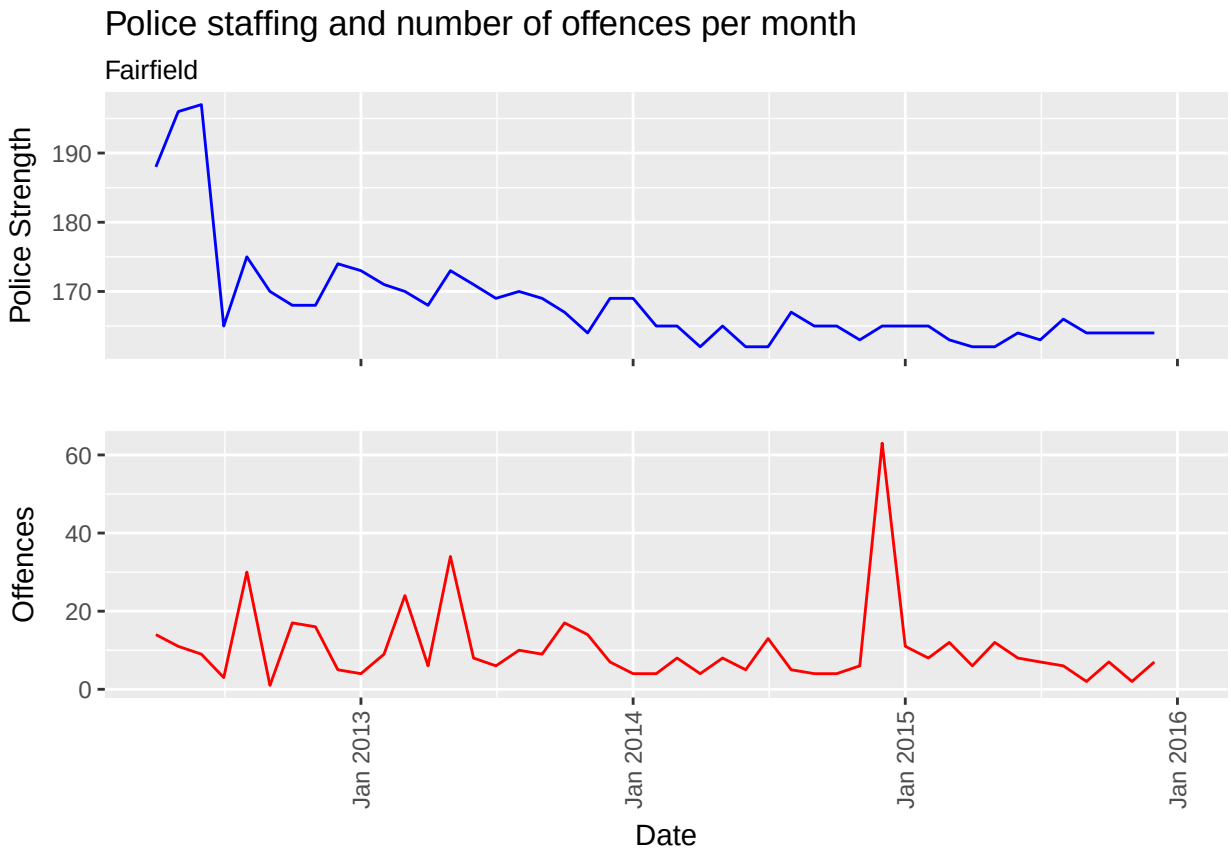

```
## [1] "LGA:  Fairfield  Crime type:  Theft"
```



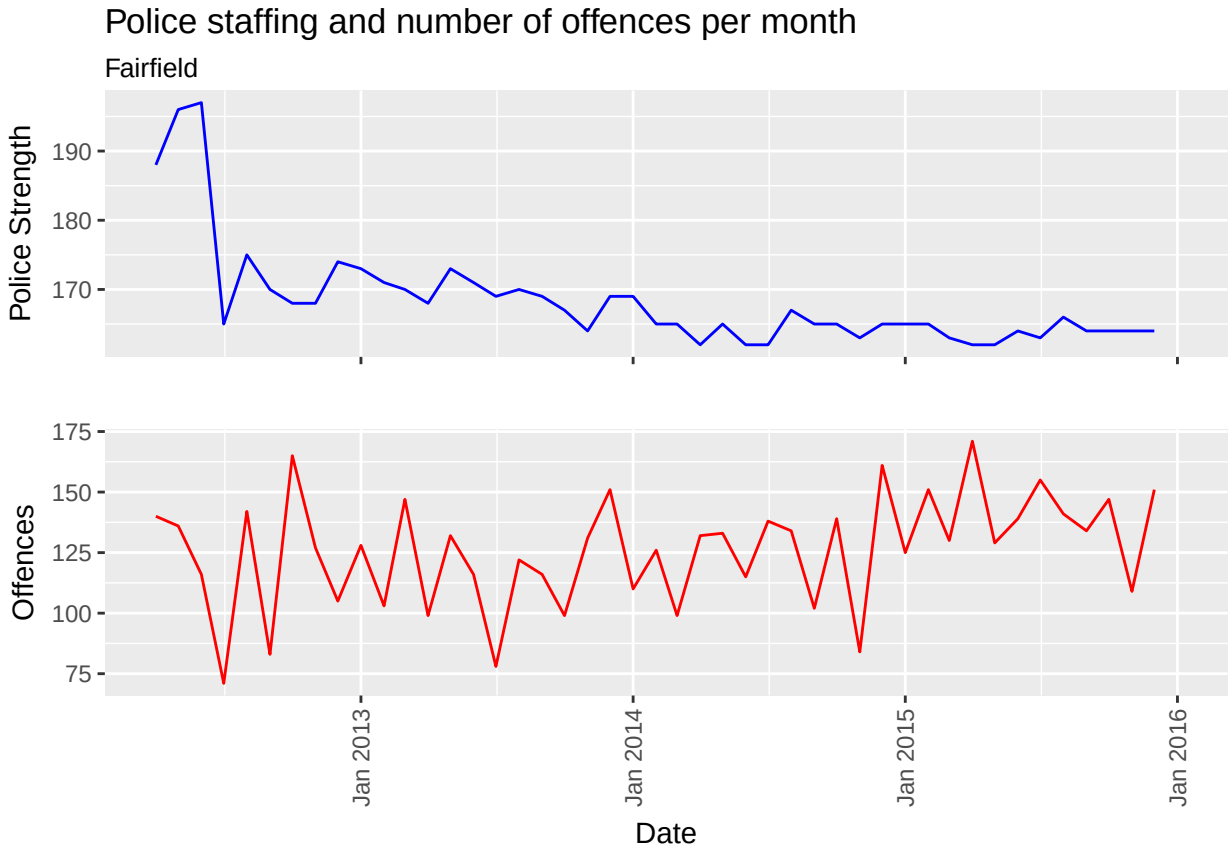

```
## [1] "LGA:  Fairfield  Crime type:  Robbery"
```



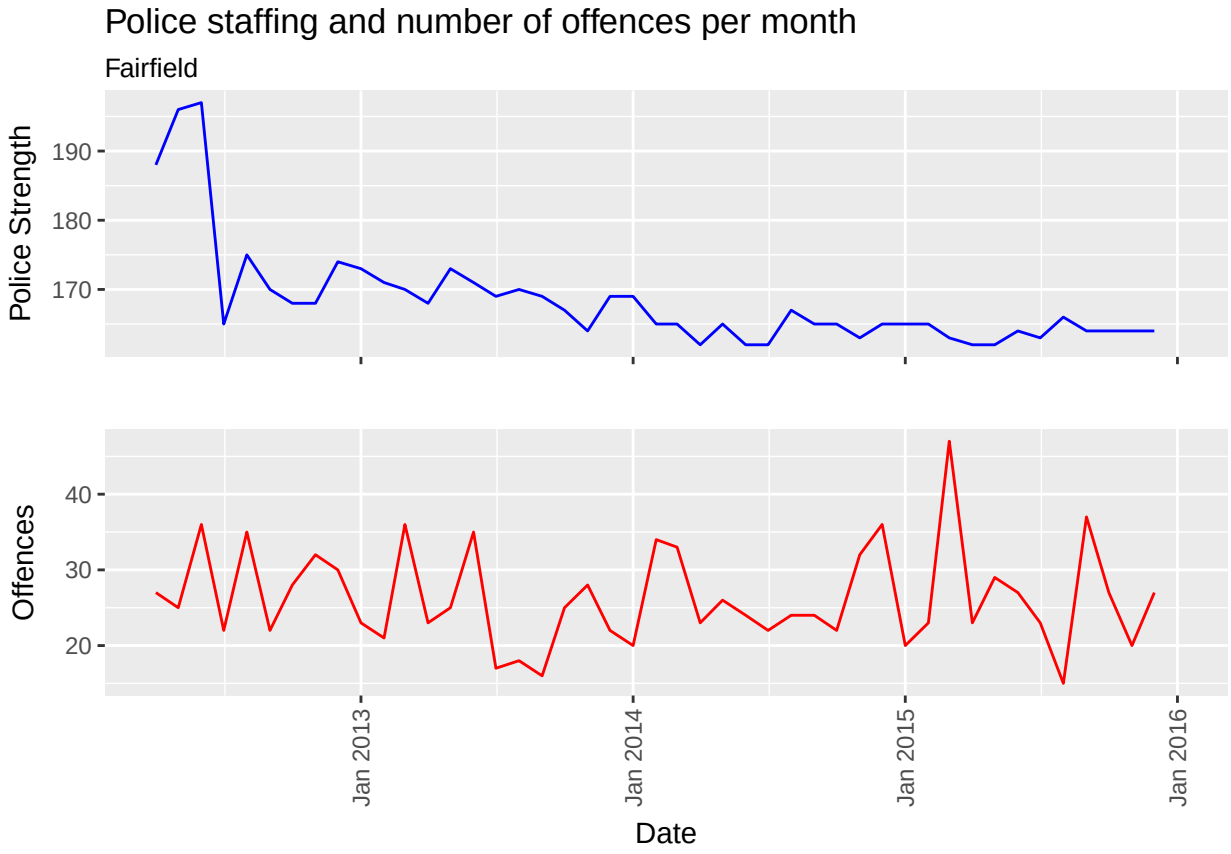

```
## [1] "LGA:  Fairfield  Crime type:  Drug Possession"
```



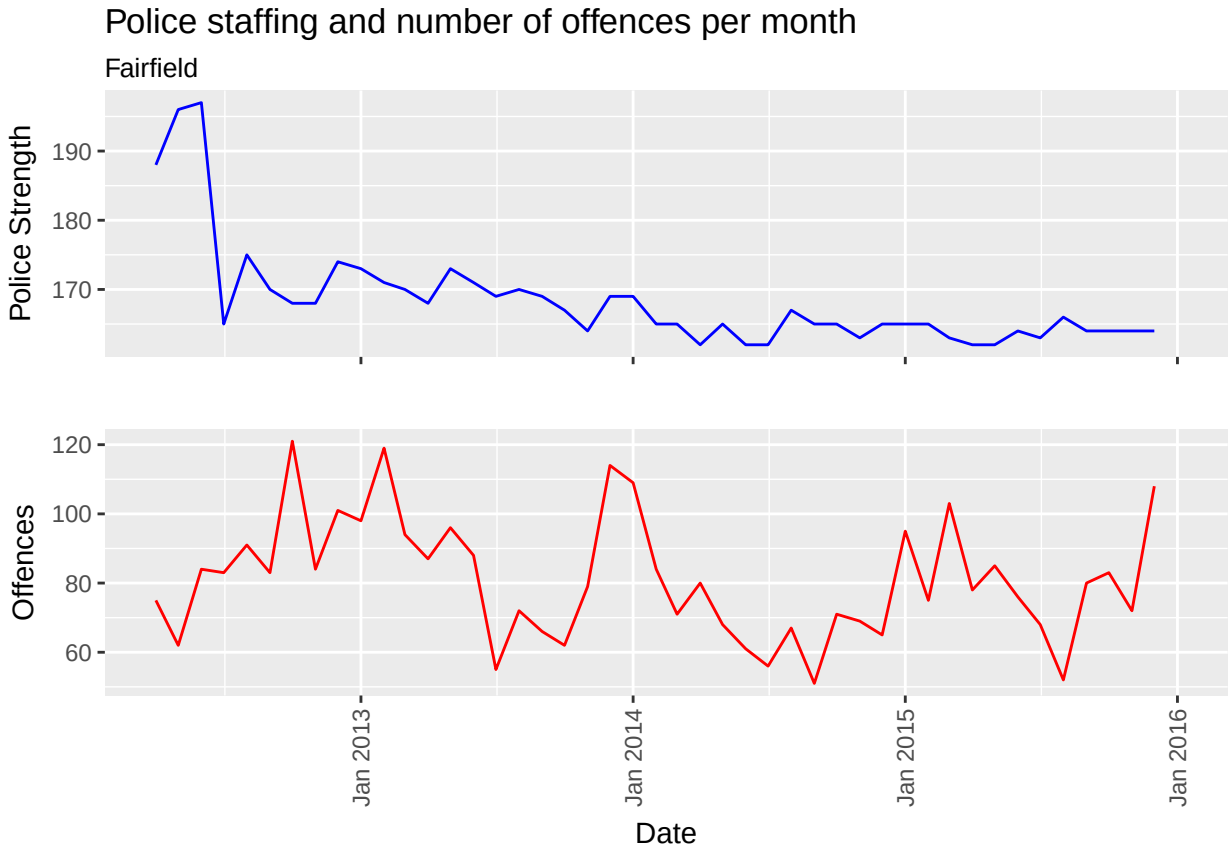

```
## [1] "LGA:  Fairfield  Crime type:  Drug Dealing"
```



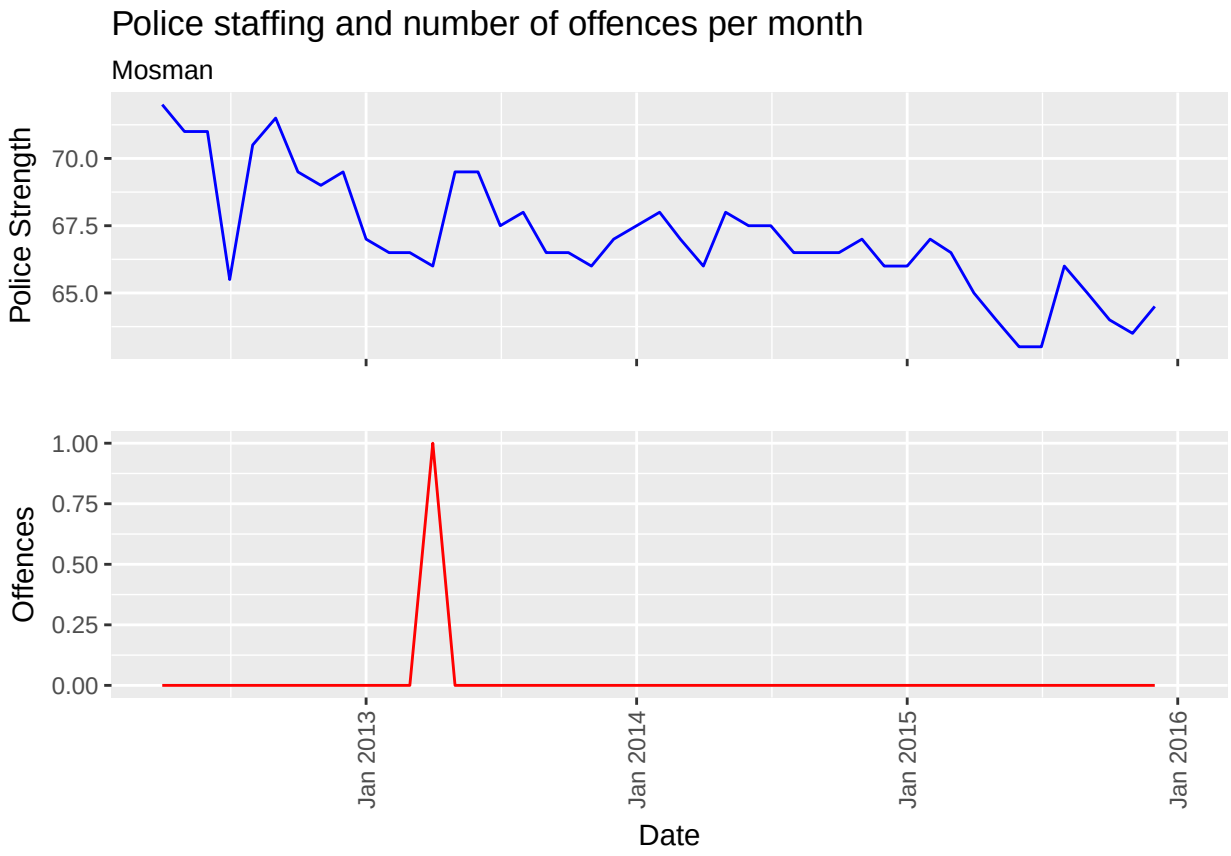

```
## [1] "LGA:  Fairfield  Crime type:  Drug Offences"
```



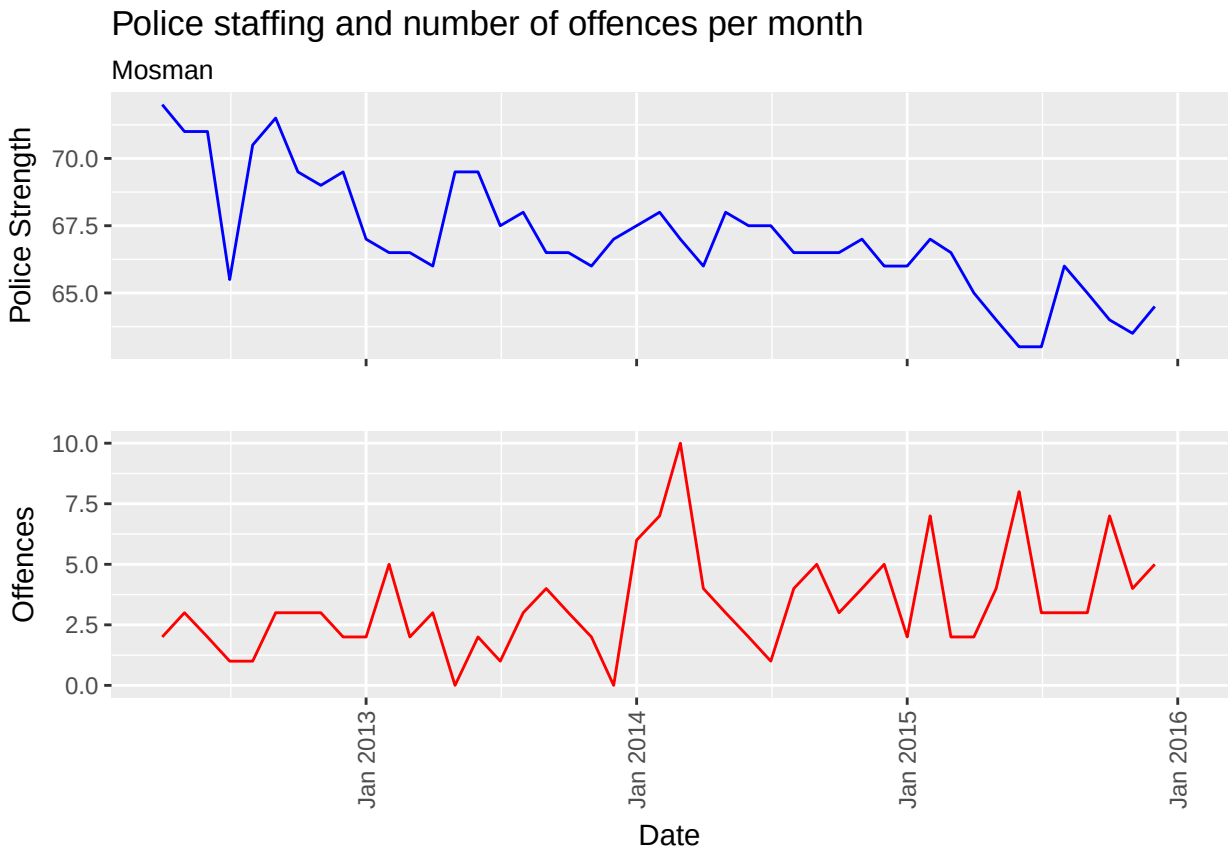

```
## [1] "LGA:  Fairfield  Crime type:  Disorderly Conduct"
```



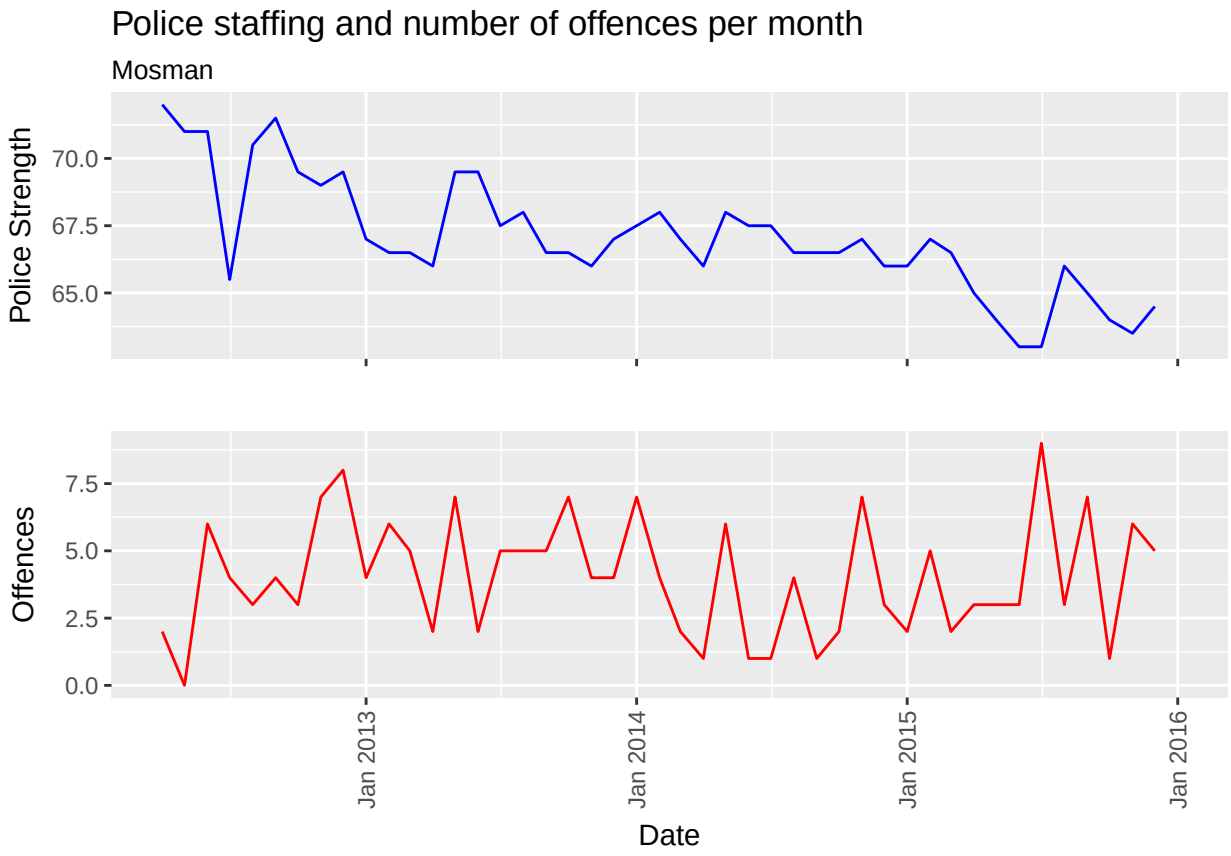

```
## [1] "LGA:  Fairfield  Crime type:  Against Justice Procedures"
```



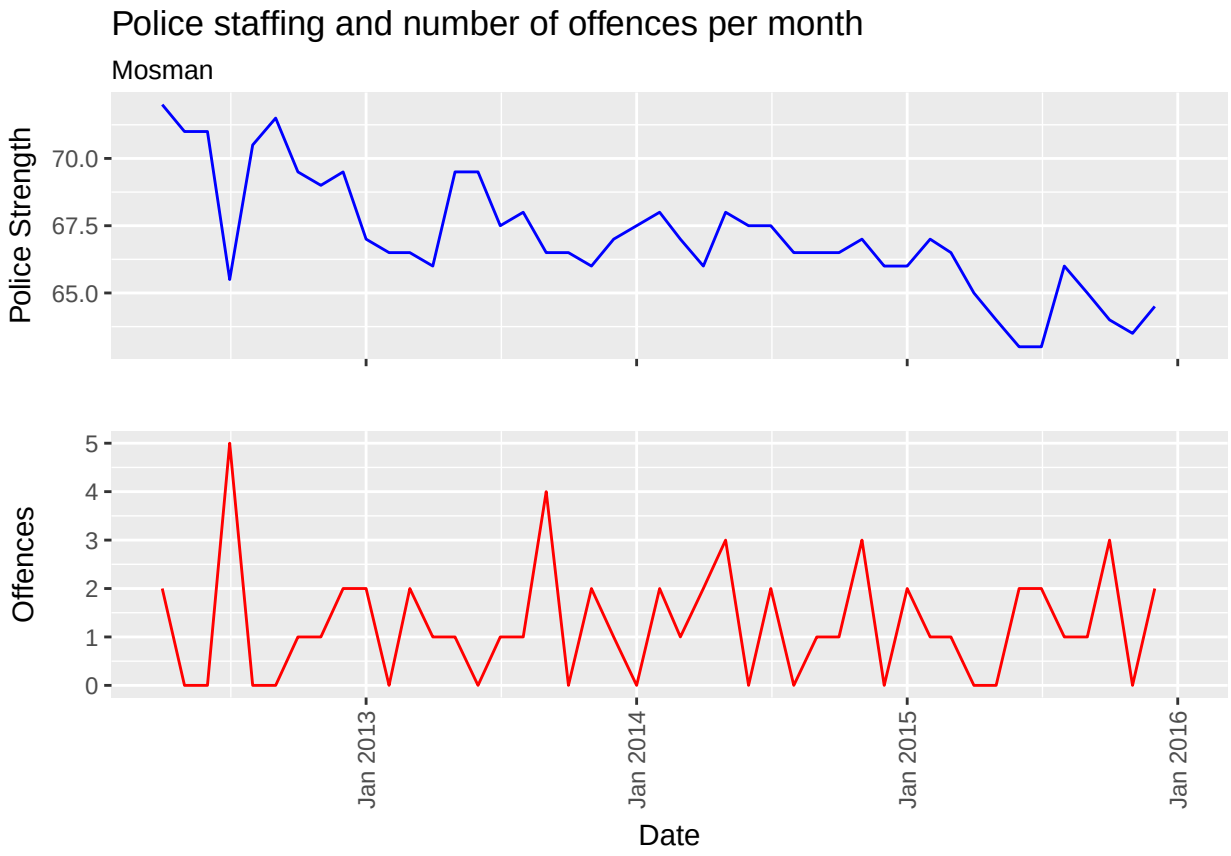

```
## [1] "LGA:  Mosman  Crime type:  Homicide"
```



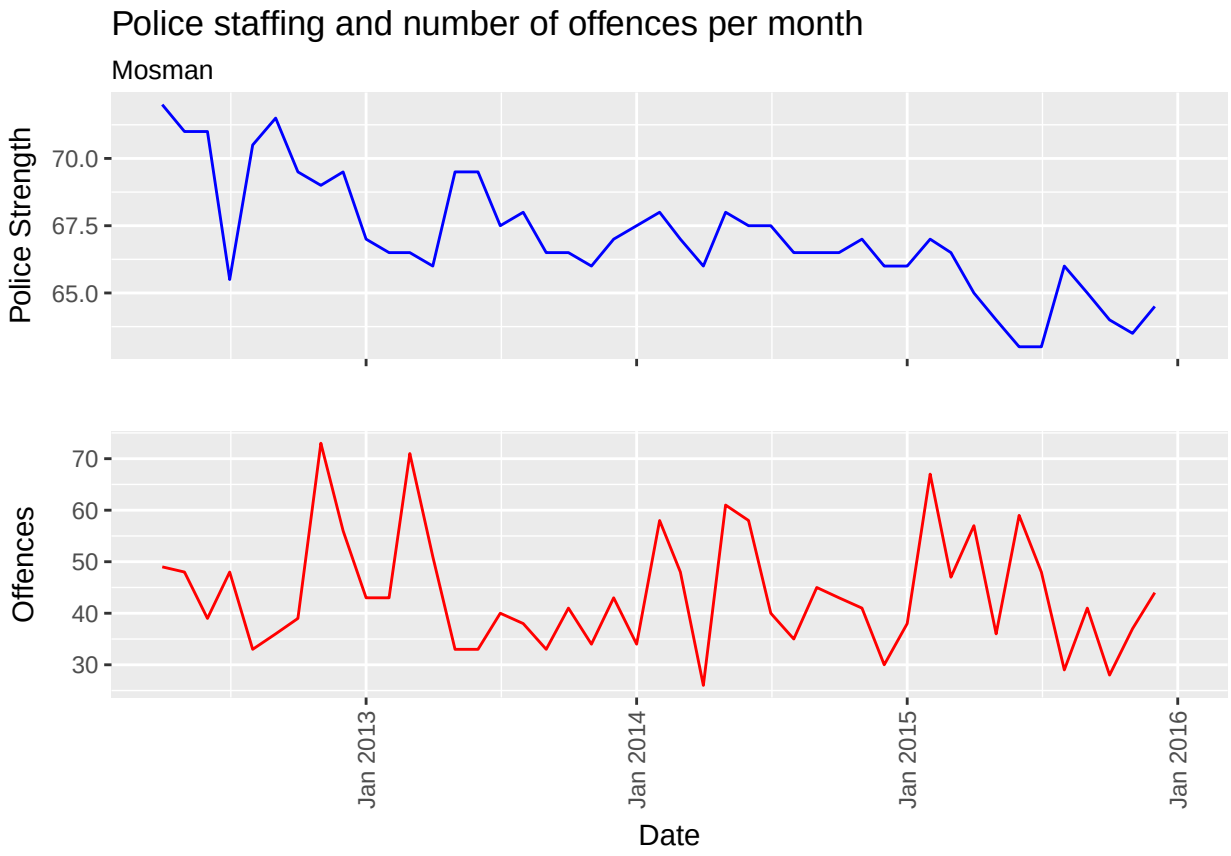

```
## [1] "LGA:  Mosman  Crime type:  Domestic Assault"
```



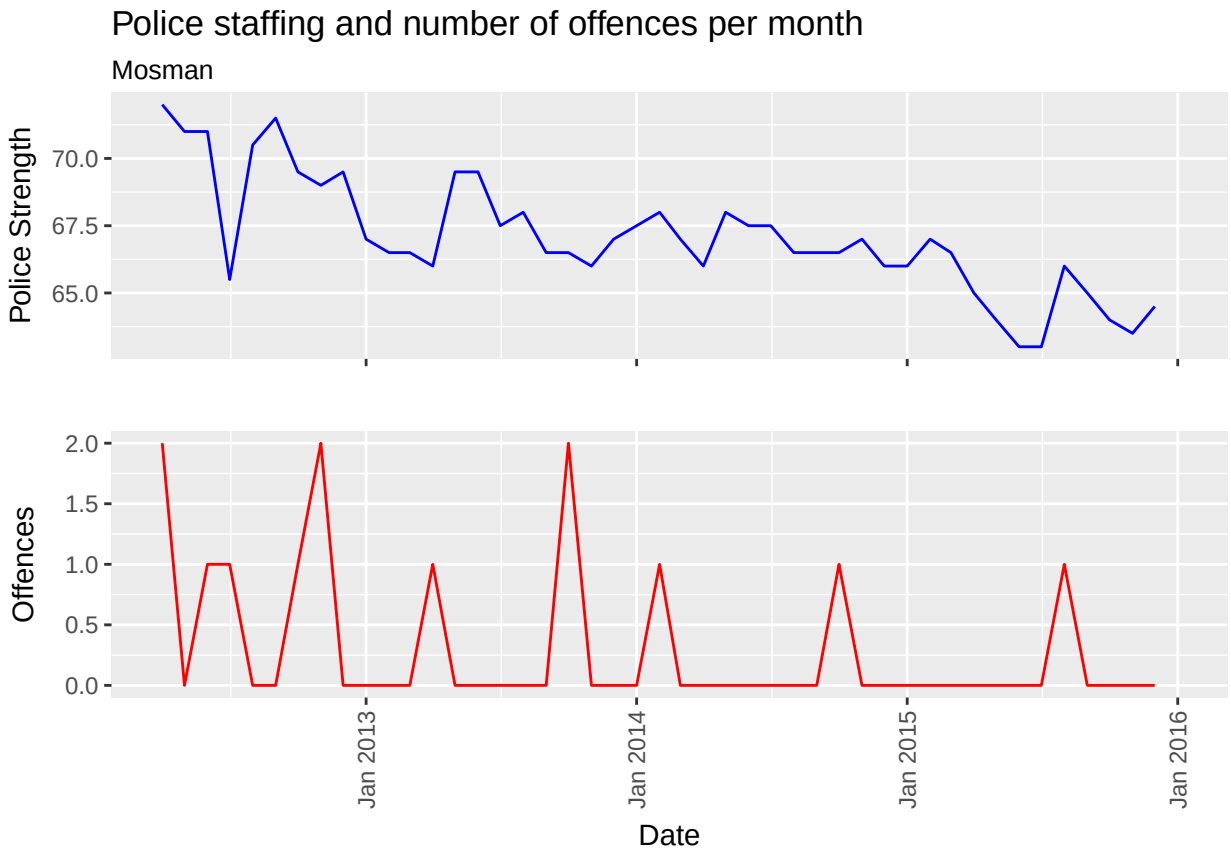

```
## [1] "LGA:  Mosman  Crime type:  Assault"
```



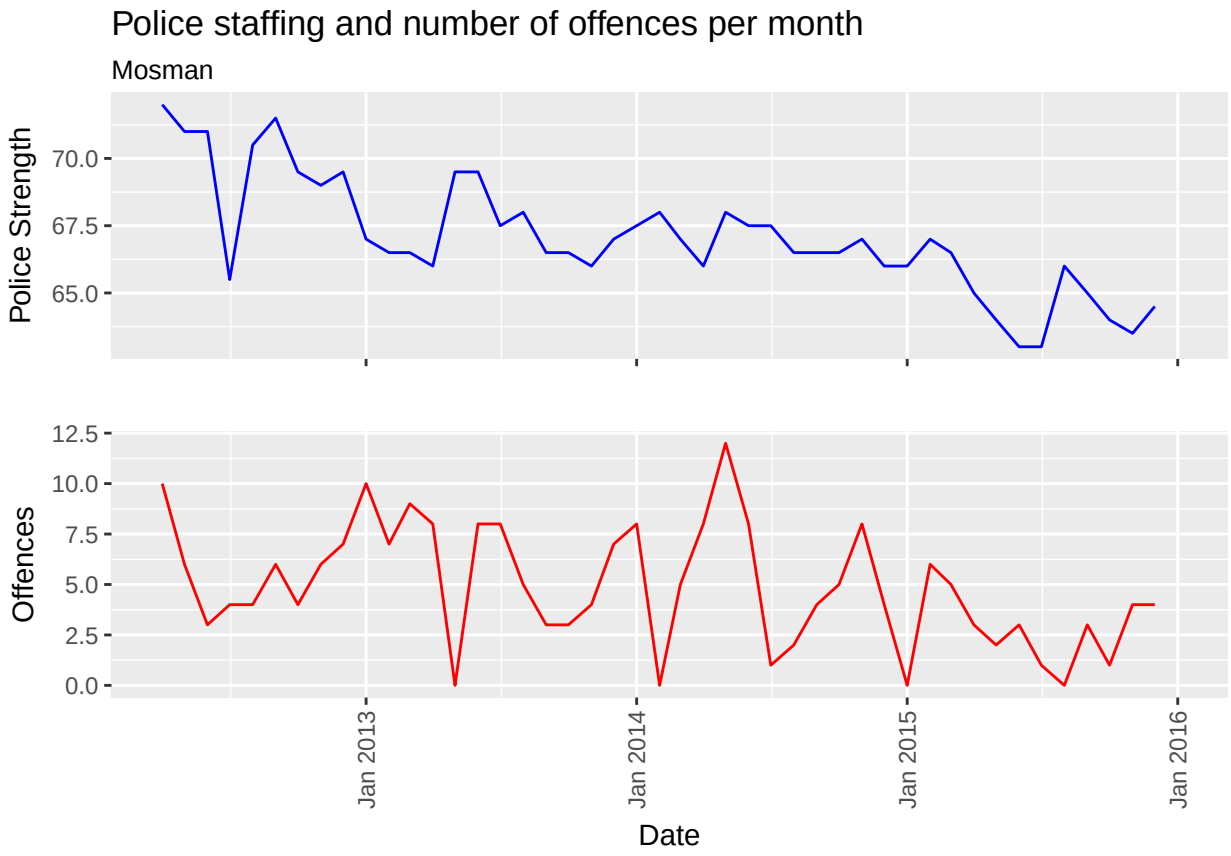

```
## [1] "LGA:  Mosman  Crime type:  Sexual Offences"
```



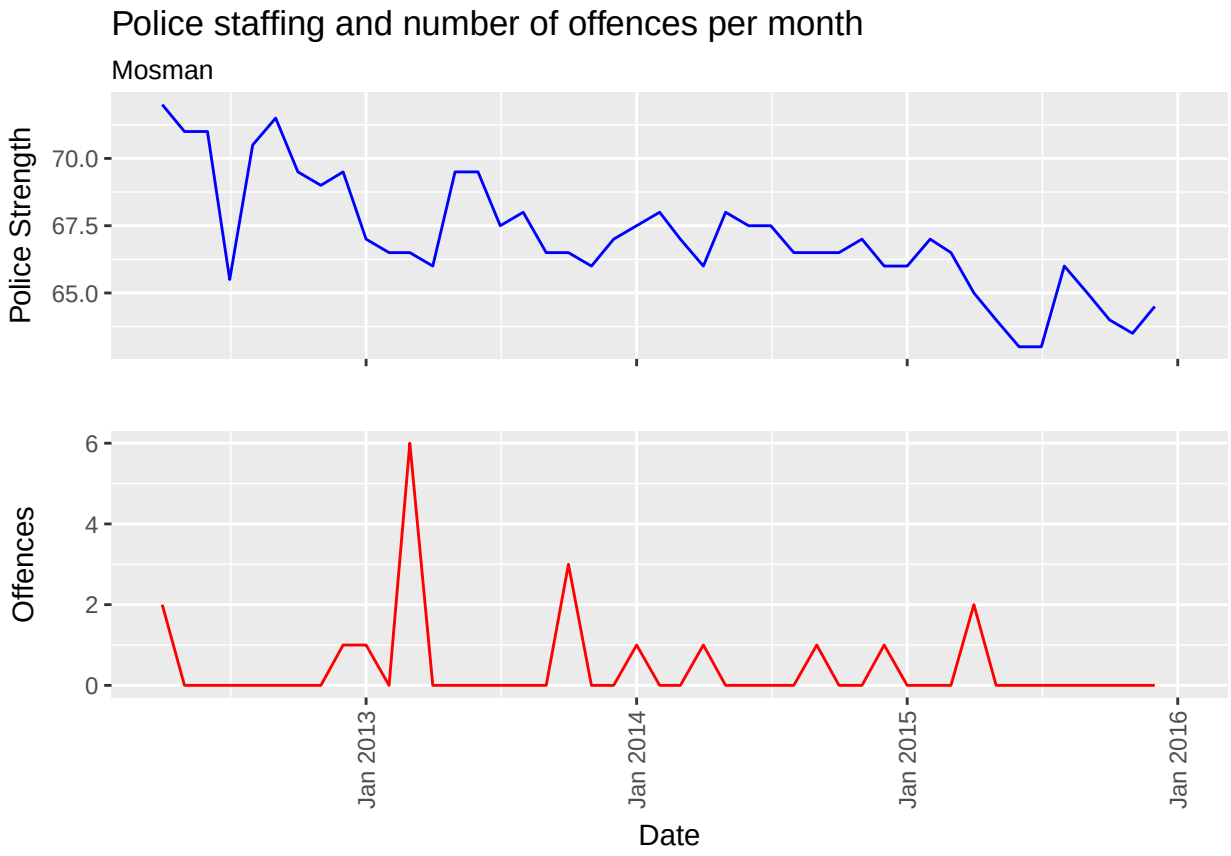

```
## [1] "LGA:  Mosman  Crime type:  Theft"
```



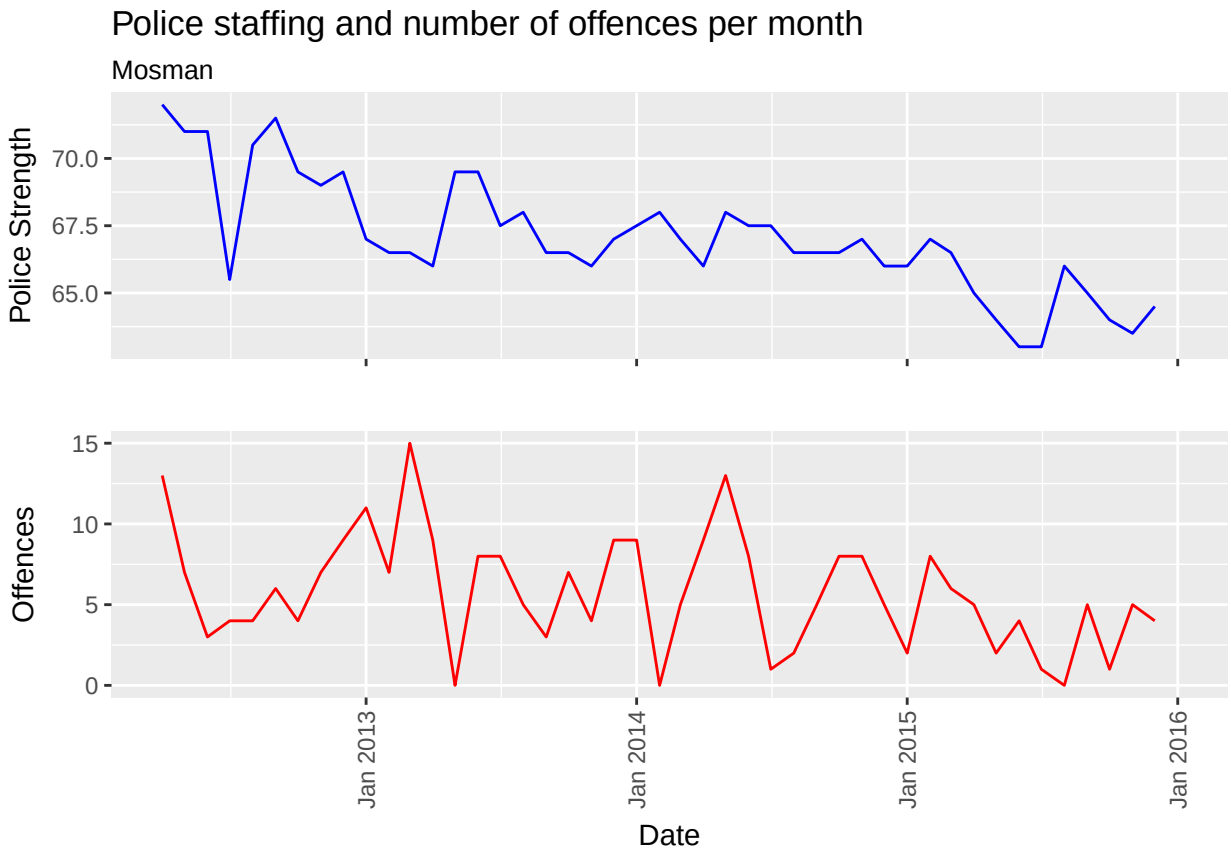

```
## [1] "LGA:  Mosman  Crime type:  Robbery"
```



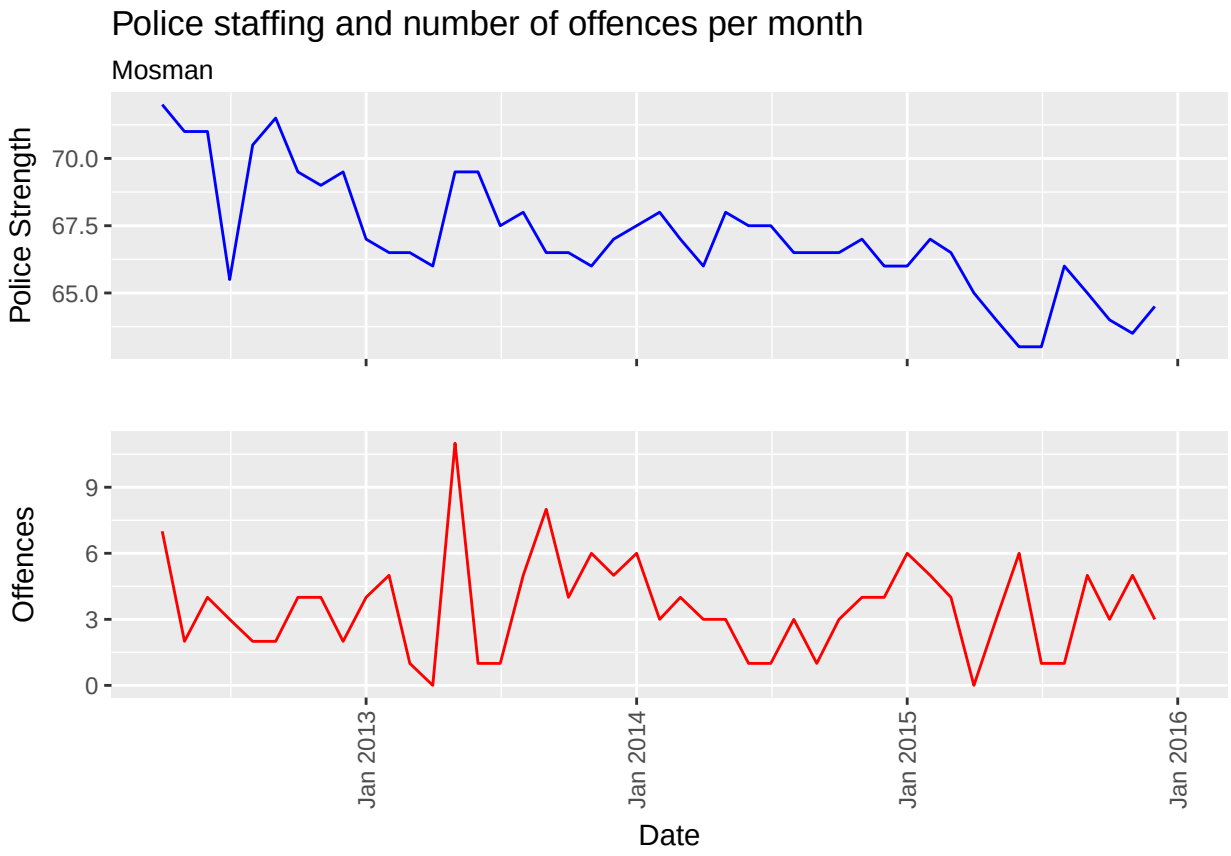

```
## [1] "LGA:  Mosman  Crime type:  Drug Possession"
```



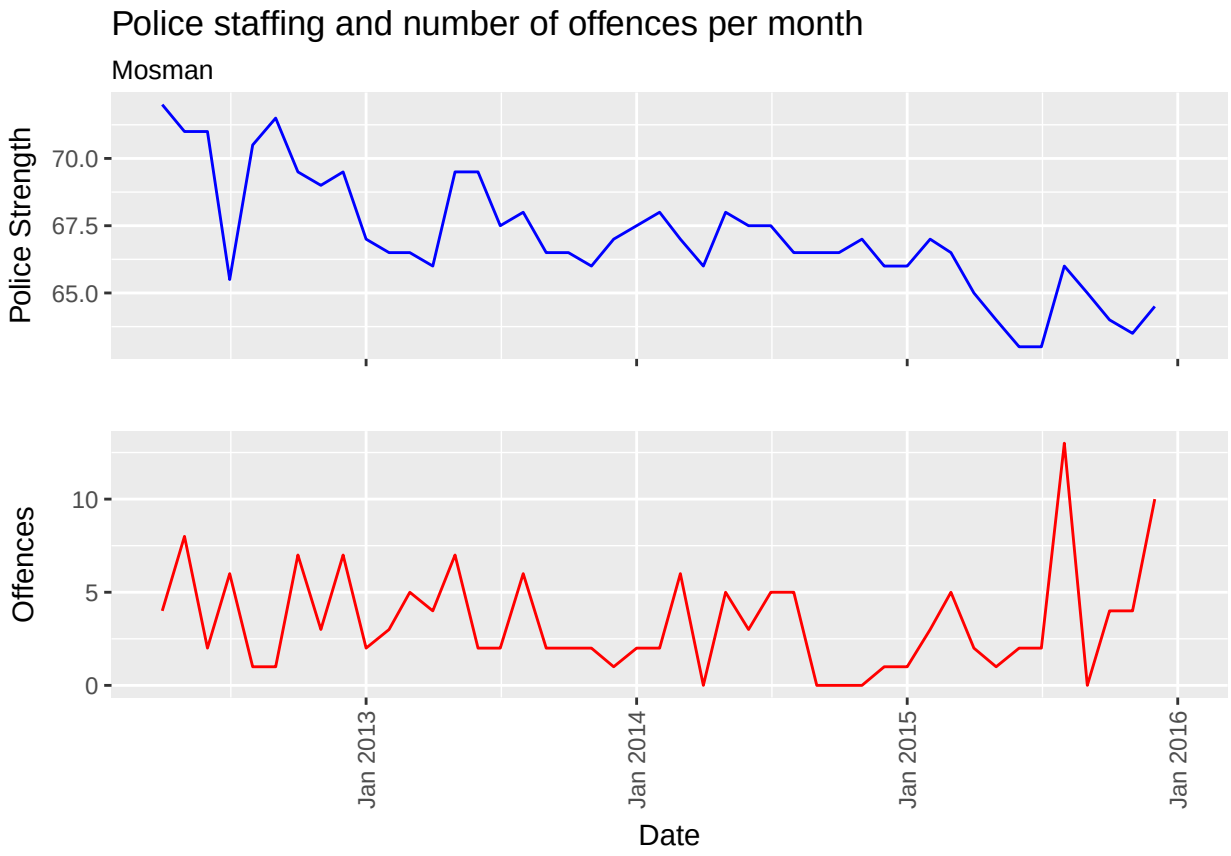

```
## [1] "LGA:  Mosman  Crime type:  Drug Dealing"
```




```
## [1] "LGA:  Mosman  Crime type:  Drug Offences"
```




```
## [1] "LGA:  Mosman  Crime type:  Disorderly Conduct"
```




```
## [1] "LGA:  Mosman  Crime type:  Against Justice Procedures"
```