# M&S plots in ggplot2

*Nathan Green, Imperial College London*

*07/09/2019*

**Slide 4**

```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
library(ggplot2)
```

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

Read in data.

```r
library(dataPakistan)

file_name <- system.file(package = "dataPakistan", "extdata", "List of AFP Cases 2015-2019.xlsx")
dat <- readxl::read_xlsx(file_name, sheet = "Data")
```

```r
##############
# preprocess #
##############

dat$month <- lubridate::month(dat$DENTER, label = TRUE)

dat$afp_cases <- dat$AFP == "ALL DIAGNOSED"

x <-
  dat %>%
  group_by(month, YRONSET) %>%
  summarise(cases = sum(afp_cases, na.rm = TRUE)) %>%
  mutate(month_year = paste(month, YRONSET)) %>%
  arrange(YRONSET, month)


#########
# plots #
#########

ggplot(x, aes(x = month_year, y = cases)) +
  geom_bar(stat = "identity")
```
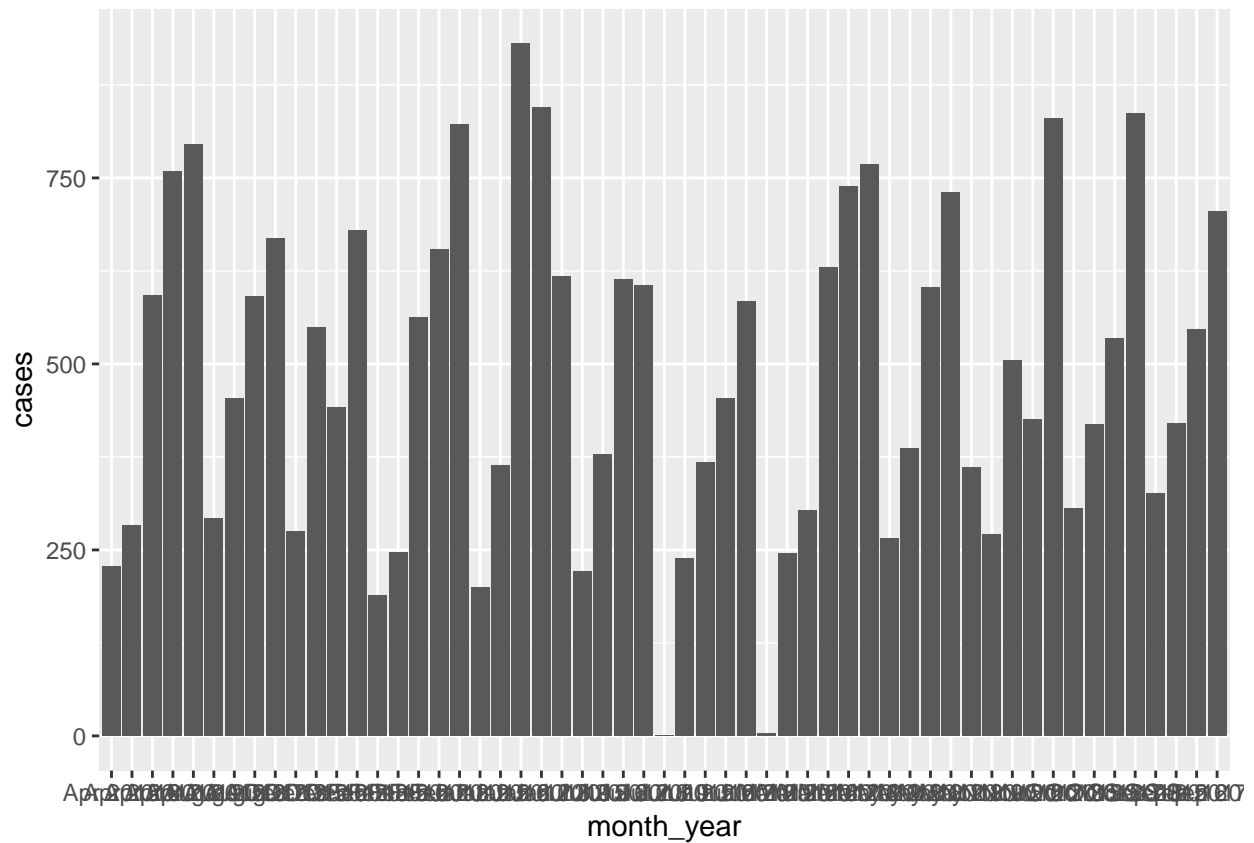
Graph 1: Distribution of AFP Cases by Month, Pakistan 2015-2019*

```r
barplot(height = x$cases, col = "darkgreen", names.arg = x$month, las = 2)
# axis(1, at = seq(nrow(x)), labels = x$month)

# plotrix::textbox(c(-10, -10), -1, textlist = "hjkl", col = "white", fill = "green")
mtext("2015", side = 1, line = 3, at = 8)
mtext("2016", side = 1, line = 3, at = 22)
mtext("2017", side = 1, line = 3, at = 36)
mtext("2018", side = 1, line = 3, at = 50)
mtext("2019", side = 1, line = 3, at = 64)

title("Graph 1:Distribution of AFP Cases by Month, Pakistan 2015-2019*
", ylab = "Cases (n)")

abline(v = 15)
abline(v = 30)
abline(v = 45)
```
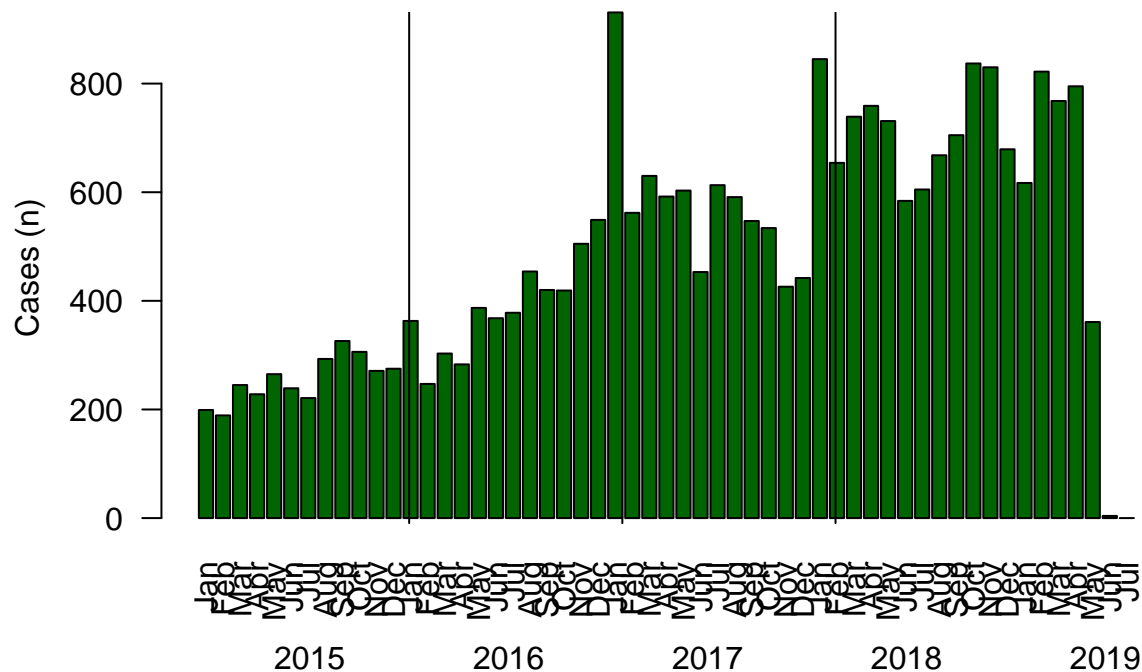
# Graph 1:Distribution of AFP Cases by Month, Pakistan 2015–2019*



```r
x$month <- factor(x$month, ordered = TRUE)
x$month_year <- factor(x$month_year, levels = x$month_year, ordered = TRUE)

ggplot(data = x, aes(x = month_year, y = cases)) +
  geom_bar(stat = "identity") +
  scale_x_discrete(breaks = x$month_year, labels = x$month,
                   expand = c(0.1,0.1)) +
  facet_grid(~ YRONSET, space = "free_x", scales = "free_x", switch = "x") +
  theme_bw() +
  theme(strip.placement = "outside",
        strip.background = element_rect(fill = NA, colour = "green"),
        panel.spacing = unit(0,"cm"),
        panel.grid.major.x = element_line(colour = NA, size = NULL, linetype = NULL,
                                          lineend = NULL, color = NULL, arrow = NULL,
                                          inherit.blank = FALSE),
        panel.grid.major.y = element_line(colour = "black", size = 1, linetype = NULL,
                                          lineend = NULL, color = NULL, arrow = NULL,
                                          inherit.blank = FALSE),
        plot.title = element_text(color = "green", size = 14, face = "bold.italic")) +
  ggtitle("Graph 1:Distribution of AFP Cases by Month, Pakistan 2015-2019*") +
  xlab("") +
  ylab("Cases (n)") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  scale_y_continuous(breaks = round(seq(min(x$cases), max(x$cases), by = 50), 1)) +
  labs(caption = "* Afp.rec Data as of  15-07-2019") +
  theme(
```

```
    plot.caption = element_text(size = 7, face = "italic")
  )
```

**Graph 1:Distribution of AFP Cases by Month, Pakistan 2015–201**
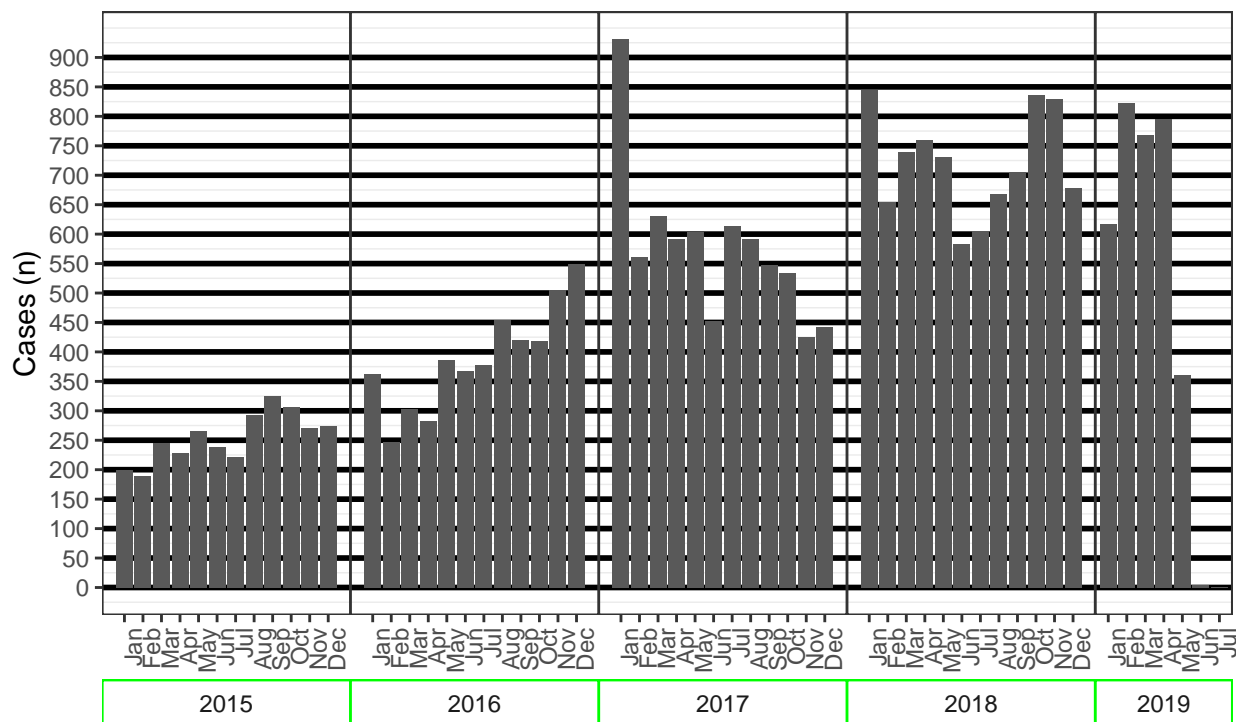


*\* Afp.rec Data as of  15–07–2019*

**Slide 5**

```
library(dplyr)
library(zoo)
```

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

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
library(reshape2)
```

```
## Warning: package 'reshape2' was built under R version 3.5.3
```

```
library(ggplot2)
library(scales)
```

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

```r
library(dataPakistan)

file_name <- system.file(package = "dataPakistan", "extdata", "List of Env Samples 2015-2019.xlsx")

# dat <- read_excel(file.choose())
dat <-
  readxl::read_xlsx(
    path = file_name,
    range = "A2:F58",
    sheet = "Sheet1") #, sheet = 1)

##############
# preprocess #
##############

dat$YRONSET <- zoo::na.locf(dat$YRONSET)

dat <-
  dat %>%
  rowwise() %>%
  mutate(check_total = sum(Positive, Negative, `Under Process`, na.rm = TRUE) == `Grand Total`)

any(!dat$check_total)
```

```
## [1] FALSE
```

```r
dat <- dat %>% select(-"Grand Total", -check_total)
dat <- dat[-nrow(dat), ]

dat$month_year <- paste(dat$MONTH, dat$YRONSET)
# dat$month <- factor(dat$month, ordered = TRUE)
dat$month_year <- factor(dat$month_year, levels = dat$month_year, ordered = TRUE)

x <- melt(dat, id.vars = c("YRONSET", "MONTH", "month_year"))
x$variable <- factor(x$variable,
                     levels = c("Positive", "Negative", "Under Process"))

x$variable <- relevel(x$variable, 'Negative')


#########
# plots #
#########

#TODO: change colour scheme

ggplot(x, aes(x = month_year, y = value, fill = variable)) +
  scale_fill_manual("legend", values = c("Negative" = "lightgreen", "Positive" = "red", "Under Process"
  geom_bar(position = "fill", stat = "identity") +
  scale_y_continuous(labels = percent_format(),
                     breaks = seq(from = 0, to = 1, by = 0.1)) +
  scale_x_discrete(breaks = x$month_year, labels = x$MONTH,
                   expand = c(0.1,0.1)) +
  facet_grid(~ YRONSET, space = "free_x", scales = "free_x", switch = "x") +
```
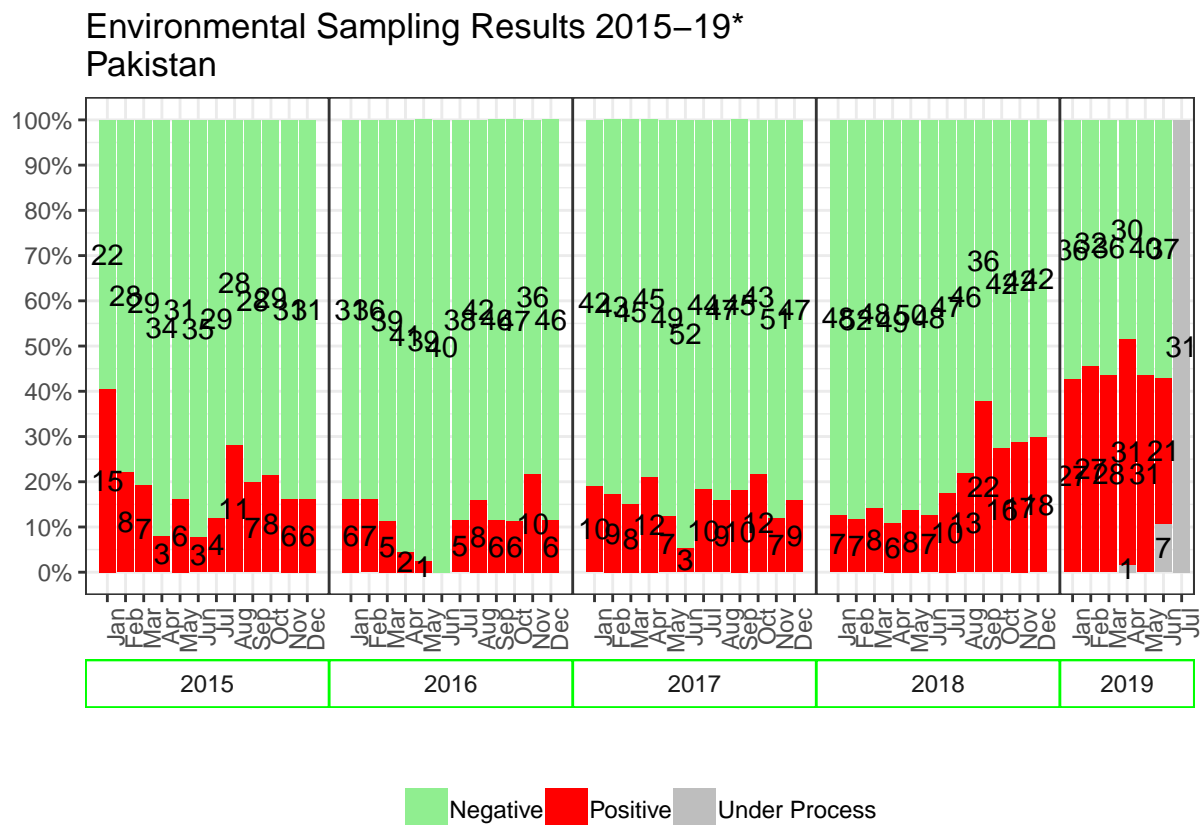
```r
  theme_bw() +
  theme(strip.placement = "outside",
        strip.background = element_rect(fill = NA, colour = "green"),
        panel.spacing = unit(0,"cm")) +
  ggtitle("Environmental Sampling Results 2015-19* \nPakistan") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  xlab("") + ylab("") +
  geom_text(data = x, aes(y = value, label = value), position = position_fill(vjust = 0.5)) +
  theme(legend.position = "bottom", legend.title = element_blank())
```

```
## Warning: Removed 55 rows containing missing values (position_stack).
```

```
## Warning: Removed 55 rows containing missing values (position_stack).
```



**Slide 7**

```r
library(dplyr)
library(zoo)
library(reshape2)
library(ggplot2)
library(scales)

library(dataPakistan)

file_name <- system.file(package = "dataPakistan", "extdata", "MPQA _March SNID.xlsx")
```

```r
dat <- readxl::read_xlsx(file_name)


##############
# preprocess #
##############

x <- dat %>% select(-DESK, -FIELD)

x$prov_district <- paste(x$"Province name", x$"District name")


#########
# plots #
#########

#TODO: change colour scheme

ggplot(x, aes(x = prov_district, y = Overall)) +
  geom_bar(stat = "identity", fill = "lightgreen") +
  geom_text(aes(label = sprintf("%d%%", Overall)), vjust = 0, angle = -90, nudge_y = -5) +
  scale_x_discrete(breaks = x$prov_district, labels = x$"District name",
                   expand = c(0.1,0.1)) +
  facet_grid(~ `Province name`, space = "free_x", scales = "free_x", switch = "x") +
  theme_bw() +
  theme(strip.placement = "outside",
        strip.background = element_rect(fill = NA, linetype = 0),
        plot.title = element_text(hjust = 0.5, face = "bold"),
        panel.spacing = unit(0,"cm")) +
  labs(subtitle =
         "- Micro Plans of Tier-1 Districts from Karachi town, Quetta Block and Peshawar/Khyber are pass
  ggtitle("SUMMARY OF MPQA RESULTS", ) +
  xlab("") + ylab("") + theme(axis.text.x = element_text(angle = 90, hjust = 1))
```
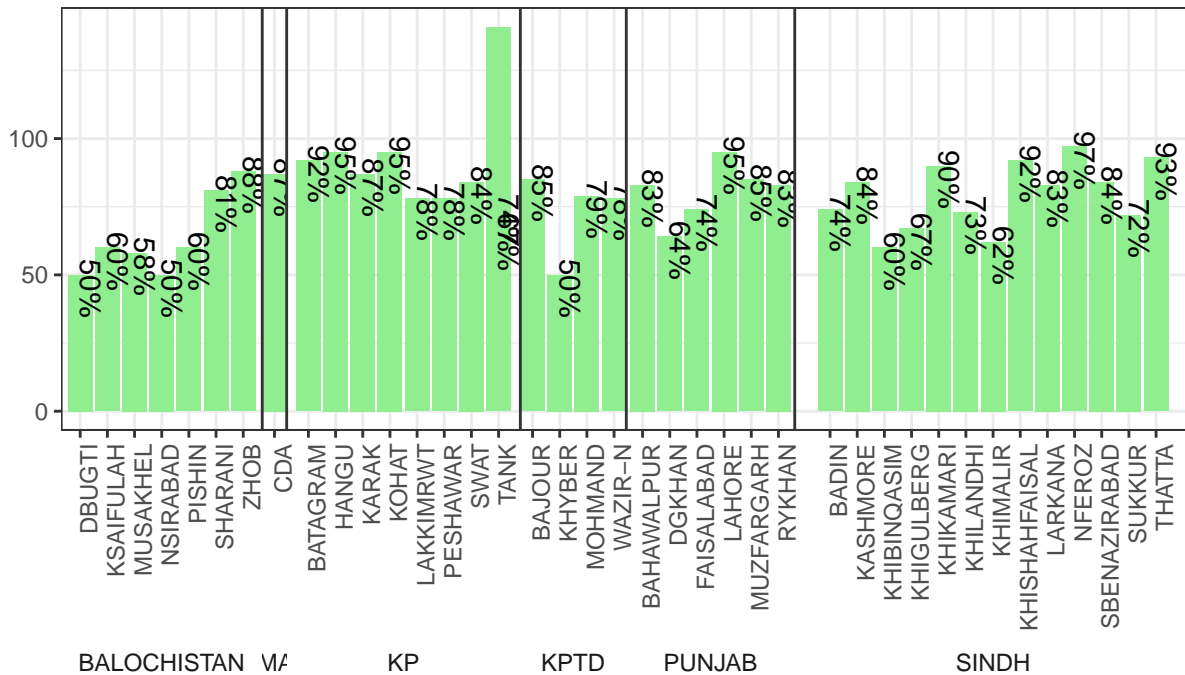
# SUMMARY OF MPQA RESULTS

– Micro Plans of Tier–1 Districts from Karachi town, Quetta Block and Peshawar/Khyber
– Substantial gaps identified primarily in Balochistan and pockets of KP and Sindh

**BALOCHISTAN:** DBUGTI 50%, KSAIFULAH 60%, MUSAKHEL 58%, NSIRABAD 50%, PISHIN 60%, SHARANI 81%, ZHOB 88%

**CDA:** 87%

**KP:** BATAGRAM 92%, HANGU 95%, KARAK 87%, KOHAT 95%, LAKKIMRWT 78%, PESHAWAR 78%, SWAT 84%, TANK 76% 67%

**KPTD:** BAJOUR 85%, KHYBER 50%, MOHMAND 79%, WAZIR–N 78%

**PUNJAB:** BAHAWALPUR 83%, DGKHAN 64%, FAISALABAD 74%, LAHORE 95%, MUZFARGARH 85%, RYKHAN 83%

**SINDH:** BADIN 74%, KASHMORE 84%, KHIBINQASIM 60%, KHIGULBERG 67%, KHIKAMARI 90%, KHILANDHI 73%, KHIMALIR 62%, KHISHAHFAISAL 92%, LARKANA 83%, NFEROZ 97%, SBENAZIRABAD 84%, SUKKUR 72%, THATTA 93%

```
# geom_text(data = x, aes(y = value, label = value), position = position_fill(vjust = 0.5))
```

**Slide 9**

```r
library(dplyr)
library(zoo)
library(reshape2)
library(ggplot2)
library(scales)

library(dataPakistan)

file_name <- system.file(package = "dataPakistan", "extdata", "MPQA _March SNID.xlsx")
dat <- readxl::read_xlsx(file_name)


##############
# preprocess #
##############

x <- dat %>% select(-DESK, -FIELD)

x$prov_district <- paste(x$"Province name", x$"District name")
```
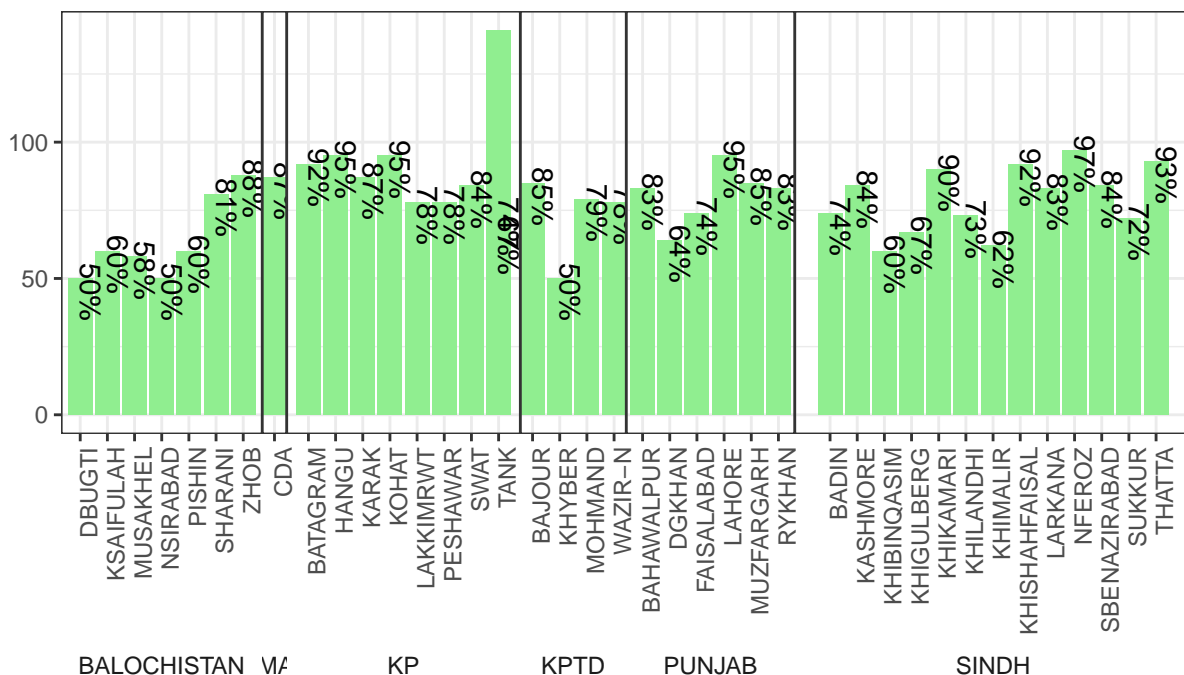
```r
#########
# plots #
#########

#TODO: change colour scheme

ggplot(x, aes(x = prov_district, y = Overall)) +
  geom_bar(stat = "identity", fill = "lightgreen") +
  geom_text(aes(label = sprintf("%d%%", Overall)), vjust = 0, angle = -90, nudge_y = -5) +
  scale_x_discrete(breaks = x$prov_district, labels = x$"District name",
                   expand = c(0.1,0.1)) +
  facet_grid(~ `Province name`, space = "free_x", scales = "free_x", switch = "x") +
  theme_bw() +
  theme(strip.placement = "outside",
        strip.background = element_rect(fill = NA, linetype = 0),
        plot.title = element_text(hjust = 0.5, face = "bold"),
        panel.spacing = unit(0,"cm")) +
  labs(subtitle =
         "- Micro Plans of Tier-1 Districts from Karachi town, Quetta Block and Peshawar/Khyber are pass
  ggtitle("SUMMARY OF MPQA RESULTS", ) +
  xlab("") + ylab("") + theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

### SUMMARY OF MPQA RESULTS

– Micro Plans of Tier–1 Districts from Karachi town, Quetta Block and Peshawar/Khyber
– Substantial gaps identified primarily in Balochistan and pockets of KP and Sindh



```r
# geom_text(data = x, aes(y = value, label = value), position = position_fill(vjust = 0.5))
```