

Your Document Title

Author Name

14 April, 2024

R Markdown

```
library(conflicted)
library(ggplot2)
library(GGally)
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg      ggplot2
```

```
library(ggpubr)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0
## v dplyr      1.1.3      v readr      2.1.4
## v forcats    1.0.0      v stringr    1.5.0
## v lubridate  1.9.3      v tibble     3.2.1
## v purrr      1.0.2      v tidyr      1.3.0
```

```
library(corrplot)
```

```
## corrplot 0.92 loaded
```

```
require(gridExtra)
```

```
## Loading required package: gridExtra
```

```
conflict_prefer("filter", "dplyr")
```

```
## [conflicted] Will prefer dplyr::filter over any other package.
```

```
conflict_prefer("lag", "dplyr")
```

```
## [conflicted] Will prefer dplyr::lag over any other package.
```

```
data <- read.csv('./data/beerhall.csv')
colnames(data) <- c("county", "region_name", "region_code", "criminals_100k",
writeLines("\nFirst few data elements:\n")
```

```
##
## First few data elements:
```

```
str(data)
```

```
## 'data.frame':    40 obs. of  7 variables:
## $ county          : chr  "Middlesex" "Surrey" "Kent" "Sussex" ...
## $ region_name     : chr  "SouthEastern" "SouthEastern" "SouthEastern" "S
## $ region_code      : int   1 1 1 1 1 1 1 1 1 1 1 ...
## $ criminals_100k   : int   200 160 160 147 178 205 183 156 173 132 ...
## $ ale_100k         : int   541 504 552 295 409 568 708 624 463 408 ...
## $ school_attend_10k: int   560 630 790 820 990 930 1020 1130 950 1090 ...
## $ church_attend_2k : int    43 48 68 67 79 69 88 97 84 97 ...
```

```
writeLines("\n")
```

```
head(data)
```

```
##      county region_name region_code criminals_100k ale_100k school_attend
## 1 Middlesex SouthEastern         1          200       541
## 2   Surrey SouthEastern         1          160       504
## 3    Kent   SouthEastern         1          160       552
## 4   Sussex SouthEastern         1          147       295
## 5    Hants SouthEastern         1          178       409
## 6    Berks SouthEastern         1          205       568
##      church_attend_2k
## 1                   43
## 2                   48
## 3                   68
## 4                   67
## 5                   79
## 6                   69
```

```
writeLines("\n")
```

```
# returns minimum, maximum, 1st quartile, median, mean and 3rd
#quartile of every variable
```

```
writeLines("Data summary:\n")
```

```
## Data summary:
```

```
summary(data)
```

```
##      county          region_name      region_code  criminals_100k
## Length:40      Length:40      Min.    :1.00      Min.    : 66.0
## Class :character Class :character 1st Qu.:1.00      1st Qu.:127.0
## Mode  :character Mode  :character Median  :3.00      Median :157.5
##                                     Mean   :3.45      Mean   :152.9
##                                     3rd Qu.:5.00      3rd Qu.:174.2
##                                     Max.    :8.00      Max.    :241.0
```

```
##      ale_100k      school_attend_10k church_attend_2k
##  Min.   : 87.0    Min.   : 560.0    Min.   : 43.0
## 1st Qu.:209.0    1st Qu.: 880.0    1st Qu.: 65.0
## Median :407.0    Median : 965.0    Median : 79.5
## Mean   :374.9    Mean   : 957.8    Mean   : 77.5
## 3rd Qu.:490.8    3rd Qu.:1082.5    3rd Qu.: 91.0
## Max.   :708.0    Max.   :1250.0    Max.   :113.0
```

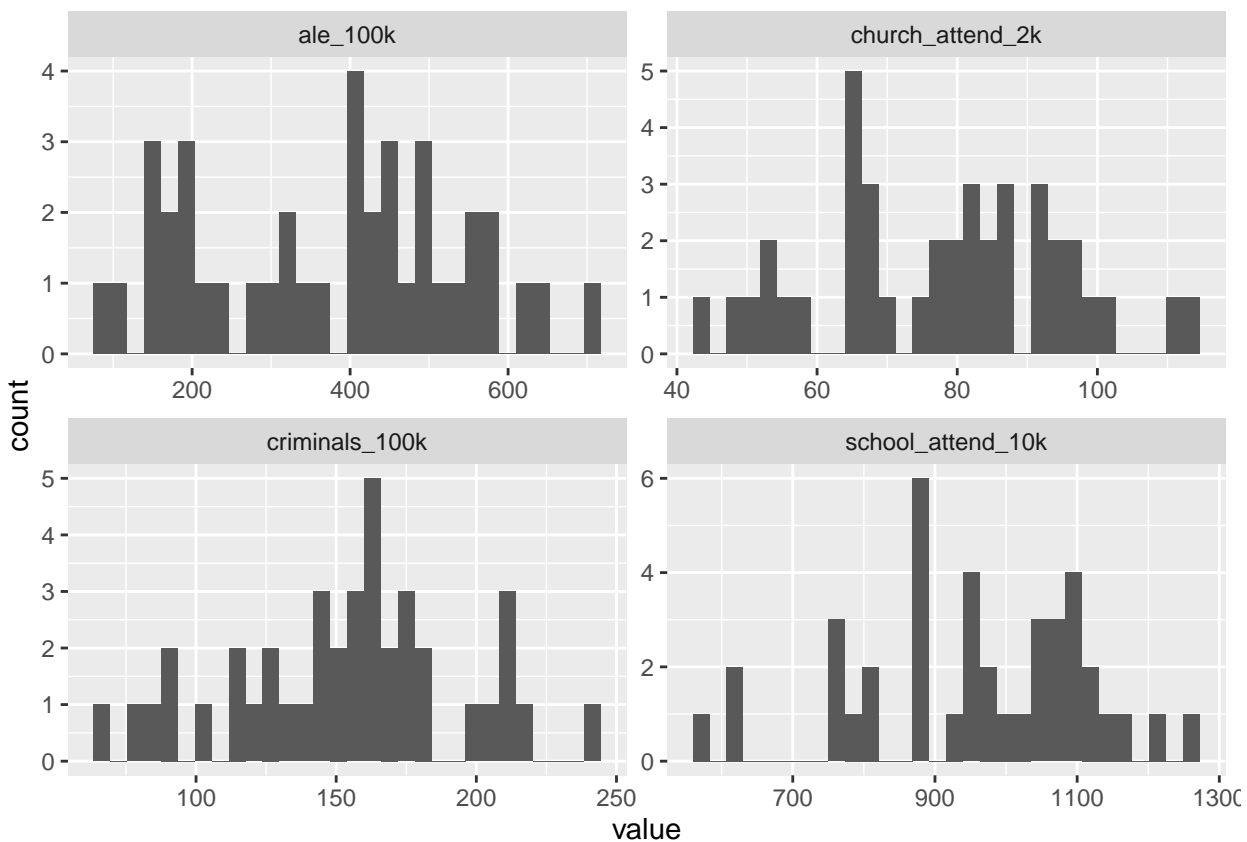
```
writeLines("\n")
```

```
# plot variable-wise histograms
```

```
numeric_data <- data[, sapply(data, is.numeric)]
```

```
data %>% pivot_longer(cols = c(criminals_100k, ale_100k, school_attend_10k, church_attend_2k)) %>%
  ggplot(aes(value)) + facet_wrap(~ name, scales = "free") + geom_histogram()
```

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

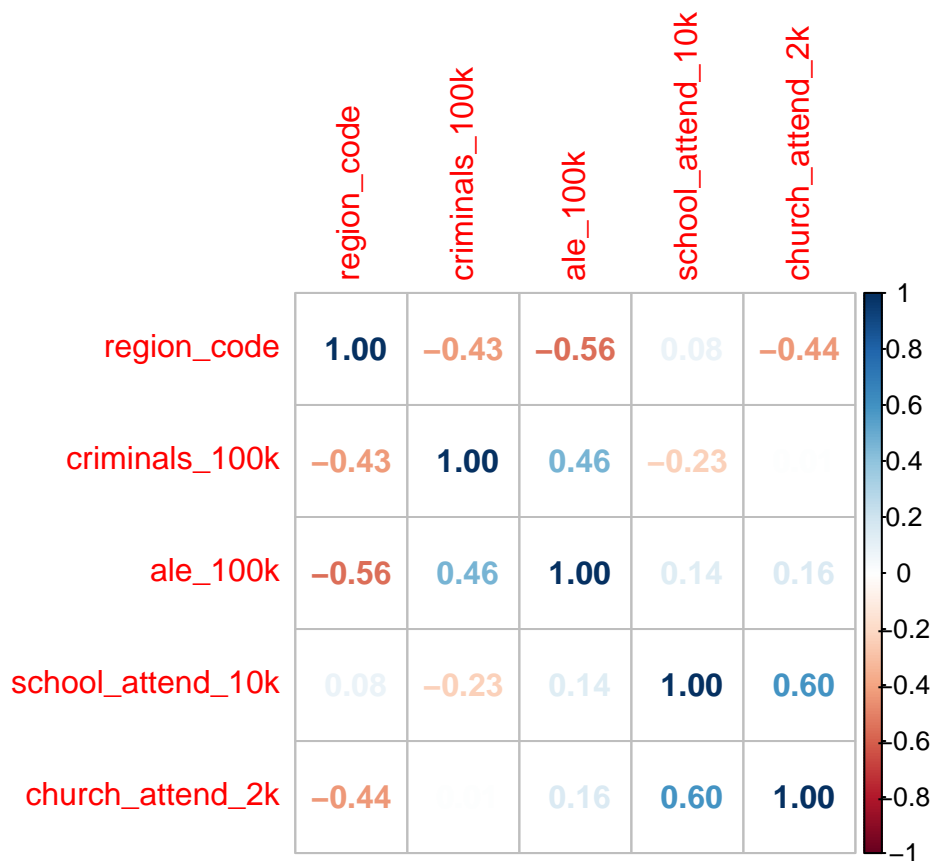


```
# analyse correlation between variables
```

```
par(mfrow = c(1, 1))
```

```
M <- cor(numeric_data)
```

```
corrplot(cor(numeric_data), method = "number")
```



```
# regression plots
fig1 <- ggplot(data, aes(ale_100k, criminals_100k)) + geom_point() +
  stat_smooth(method = "lm")
fig2 <- ggplot(data, aes(school_attend_10k, criminals_100k)) + geom_point() +
  stat_smooth(method = "lm")
fig3 <- ggplot(data, aes(region_code, criminals_100k)) + geom_point() +
  stat_smooth(method = "lm")
fig4 <- ggplot(data, aes(church_attend_2k, criminals_100k)) + geom_point() +
  stat_smooth(method = "lm")

ggarrange(fig1, fig2, fig3, fig4)
```

```
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
```

