JMP_defects3

2023-10-25

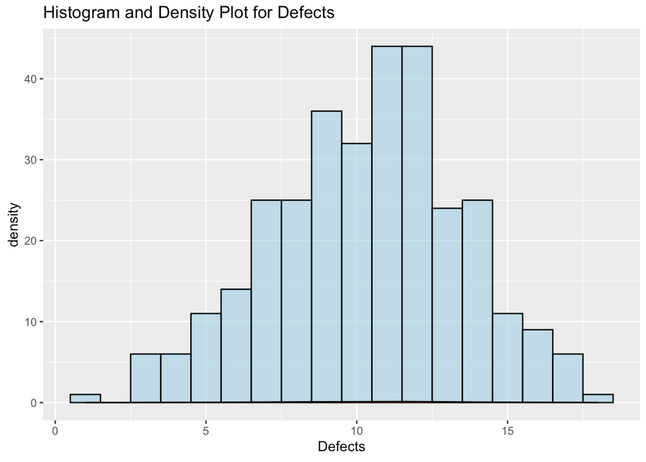
```
library(haven)
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──
## ✔ ggplot2 3.4.0      ✔ purrr   1.0.0
## ✔ tibble  3.1.8      ✔ dplyr   1.0.10
## ✔ tidyr   1.2.1      ✔ stringr 1.5.0
## ✔ readr   2.1.3      ✔ forcats 0.5.2
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
```

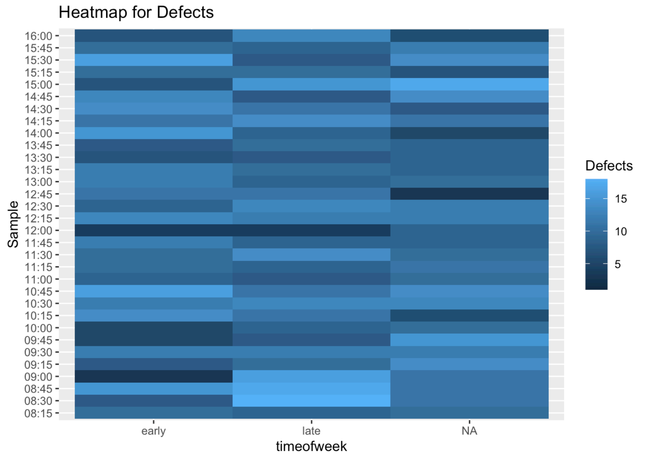
```
library(plotly)
```

```
##
## Attaching package: 'plotly'
##
## The following object is masked from 'package:ggplot2':
##
##     last_plot
##
## The following object is masked from 'package:stats':
##
##     filter
##
## The following object is masked from 'package:graphics':
##
##     layout
```

```
# Read the data from the CSV file
defects <- read.csv("defects.csv")

defects$timeofweek <- ifelse(defects$Day %in% c(1, 2, 6, 7), "early",
                        ifelse(defects$Day %in% c(4, 5, 9, 10), "late", NA))


# Summary statistics and univariate analysis
summary_stats <- summary(defects$Defects)

# Histogram and density plot
ggplot(defects, aes(x = Defects)) +
  geom_histogram(binwidth = 1, fill = "lightblue", color = "black", alpha = 0.7) +
  geom_density(alpha = 0.5, fill = "#FF6666") +
  labs(title = "Histogram and Density Plot for Defects")
```




```
# Calculate mean by timeofday
defects_summary <- defects %>%
  group_by(timeofweek) %>%
  summarize(mean_defects = mean(Defects))

# Heatmap
ggplot(defects, aes(x = timeofweek, y = Sample, fill = Defects)) +
  geom_tile() +
  labs(title = "Heatmap for Defects")
```




```
# ANOVA for defects
model <- glm(Day ~ timeofweek, data = defects, family = gaussian)
summary(model)
```

```
##
## Call:
## glm(formula = Day ~ timeofweek, family = gaussian, data = defects)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
##  -3.000   -2.250    0.000    2.250    3.000
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      4.0000     0.2262  17.681   <2e-16 ***
## timeofweeklate   3.0000     0.3199   9.377   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 6.551181)
##
##     Null deviance: 2240  on 255  degrees of freedom
## Residual deviance: 1664  on 254  degrees of freedom
##   (64 observations deleted due to missingness)
## AIC: 1211.7
##
## Number of Fisher Scoring iterations: 2
```