

Statistical Modelling in R - Assignment 1

- Palak Nagdev, Roll number 22

```
----> library(tidyverse, MASS, readxl)
      data = read.csv(file.choose())
```

1. Code:

```
summary(data)
```

2. Code:

```
View(head(data, 6))
```

3.Code:

```
View(tail(data, 10))
```

4. Code:

```
dim(data)
```

5. Code:

```
class(data)
```

6. Code:

```
colnames(data)
```

7. Code:

```
numeric_data = data[unlist(lapply(data, FUN = is.numeric))]  
for (col in colnames(numeric_data)) {  
  print(col)  
  print(fivenum(unlist(numeric_data[col])))  
}
```

8. Code:

```
data[3]
```

9. Code:

```
data_lis = data[data$License.. > 2000000 & data$License.. < 3000000,]
```

10. Code:

```
blue_line = data[data$DBA.Name == "BLUE LINE", ]
```

11. Code:

```
data[, c("Ã..Inspection.ID", "Facility.Type")]
```

12. Code:

```
data[grep("High", data$Risk), "DBA.Name"]
```

13. Code:

```
View(head(data[data$Results == "Fail",], 2))
```

14.

Code:

```
nrow(data)
```

15. Code:

```
which(unlist(lapply(data, is.na)))
```

16. Code:

```
df = data %>% select_if(~ !all(is.na(.)))
```

17. Code:

```
ggplot(data, mapping = aes(x = Latitude)) +  
geom_histogram(color = "red", fill = "black") +  
ggtitle("Latitude Histogram") +  
theme_minimal() +  
theme(plot.title = element_text(hjust = 0.5))
```

```
ggplot(data, mapping = aes(x = Longitude)) +  
geom_histogram(color = "red", fill = "black") +  
ggtitle("Longitude Histogram") +  
theme_minimal() +  
theme(plot.title = element_text(hjust = 0.5))
```

18. Code:

```
is.na(data$Latitude) = mean(data$Latitude)  
is.na(data$Longitude) = mean(data$Longitude)
```

19. Code:

```
which(unlist(lapply(data, duplicated)) == T)
```

20. Code:

```
data %>% count(Risk, Results) %>% filter(Results == "Pass")
```

21. Code:

```
data$Inspection.Date = as.Date(data$Inspection.Date, format = "%m/%d/%Y")  
data["Year"] = format(data$Inspection.Date, "%Y")  
data["Month"] = format(data$Inspection.Date, "%B")  
data["Day"] = format(data$Inspection.Date, "%d")
```