MDA102-Statistical methods u

Day 11 – 14 july 2020
6:30 AM – 7:30 AM
Dr Sharon Varghese A

dplyr package

```r
install.packages("dplyr")
library(dplyr)

Attaching package: 'dplyr'
The following objects are masked from 'pacl

    filter, lag
The following objects are masked from 'pacl

    intersect, setdiff, setequal, union
```

```
chicago<-readRDS("chicago.rds")
dim(chicago)
[1] 6940    8
class(chicago)
[1] "data.frame"
```

select() – to select desired variable columns

```
chicago_subset<-select(chicago,city:dp
head(chicago_subset)
  city tmpd    dptp
1 chic 31.5 31.500
2 chic 33.0 29.875
3 chic 33.0 27.375
4 chic 29.0 28.625
5 chic 32.0 28.875
6 chic 40.0 35.125
```
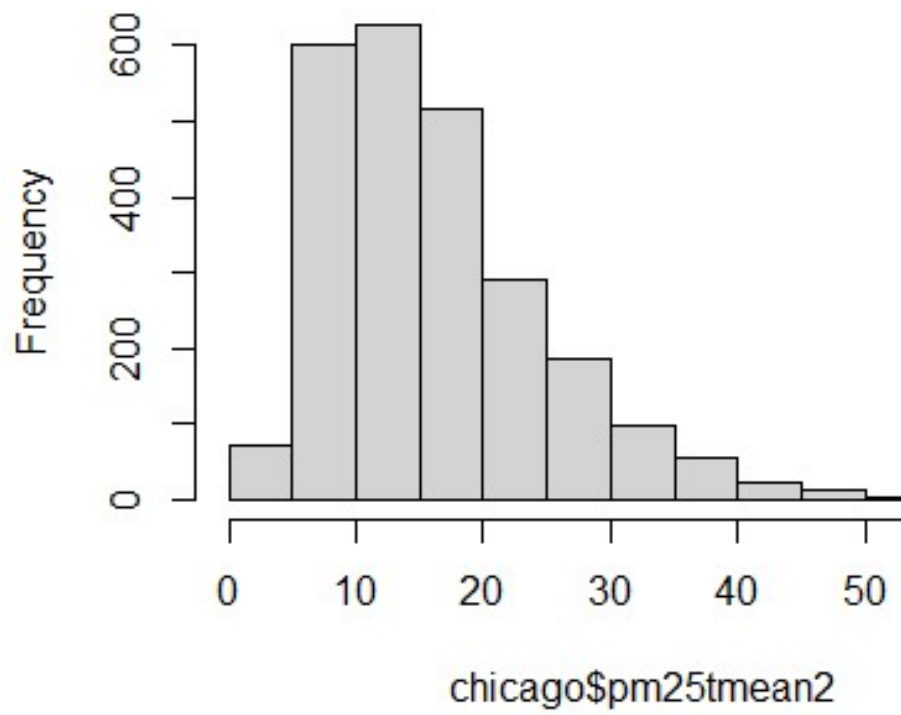
- ```
chicago_subset2<-select(chicago,-(cit
chicago_subset3<-select(chicago,ends_
variable ending with 2
chicago_subset4<-select(chicago,starts
```

**filter() – to filter out observations**
```
hist(chicago$pm25tmean2)
```

# Histogram of chicago$pm25tr



chicago$pm25tmean2

```
chicago_filtered<-filter(chicago,pm25tmean2
head(chicago_filtered)
  city tmpd dptp       date pm25tmean2 pm10
1 chic   48 46.7 1998-01-05      11.50   20
2 chic   18  7.1 1998-01-11      10.20   30
3 chic   37 30.6 1998-01-29      21.40   30
4 chic   32 25.6 1998-02-04      13.90   12
5 chic   39 33.9 1998-02-10      29.50   62
6 chic   42 35.3 1998-02-16      22.55   33
max(chicago_filtered$tmpd)
[1] 79
```