Hands on-Exercise 1

1.2.2 Importing data

```
pacman::p_load(tidyverse)

exam_data <- read_csv("data/Exam_data.csv")
```

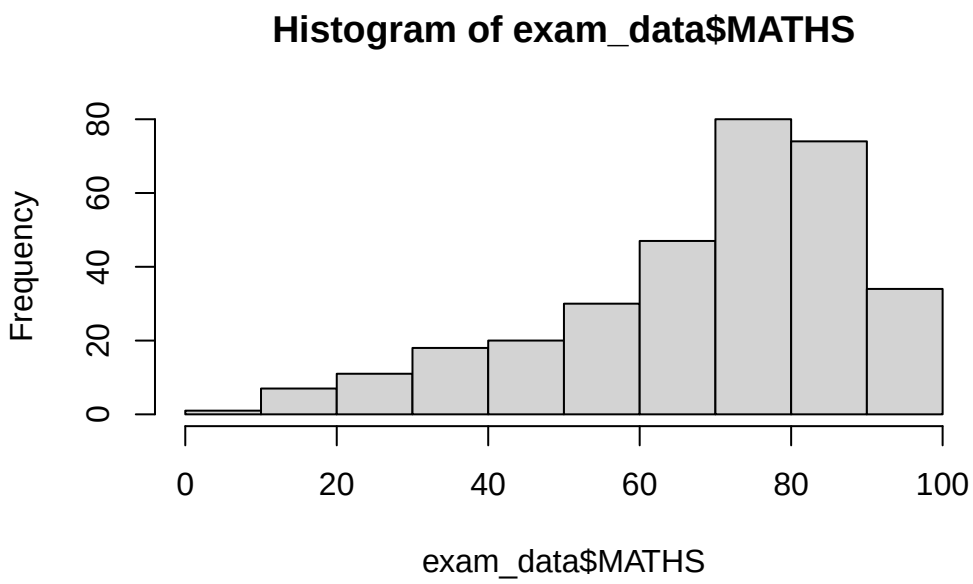
```
Rows: 322 Columns: 7
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (4): ID, CLASS, GENDER, RACE
dbl (3): ENGLISH, MATHS, SCIENCE

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

1.3.1 R Graphics VS ggplot

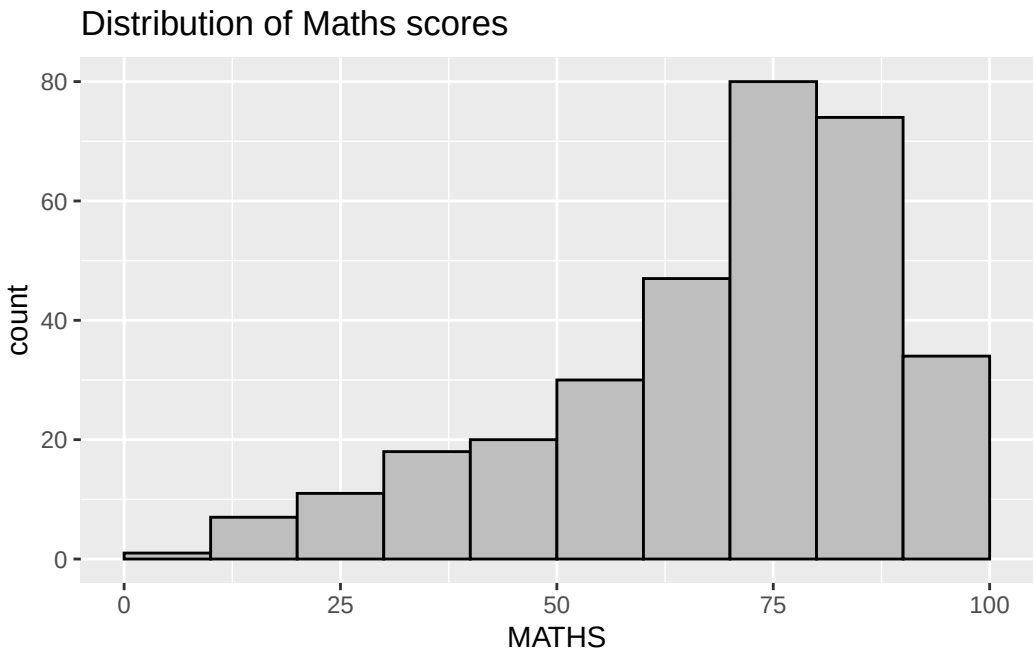
R Graphics

```
hist(exam_data$MATHS)
```



ggplot2

```
ggplot(data=exam_data, aes(x = MATHS)) +
  geom_histogram(bins=10,
                 boundary = 100,
                 color="black",
                 fill="grey") +
  ggtitle("Distribution of Maths scores")
```

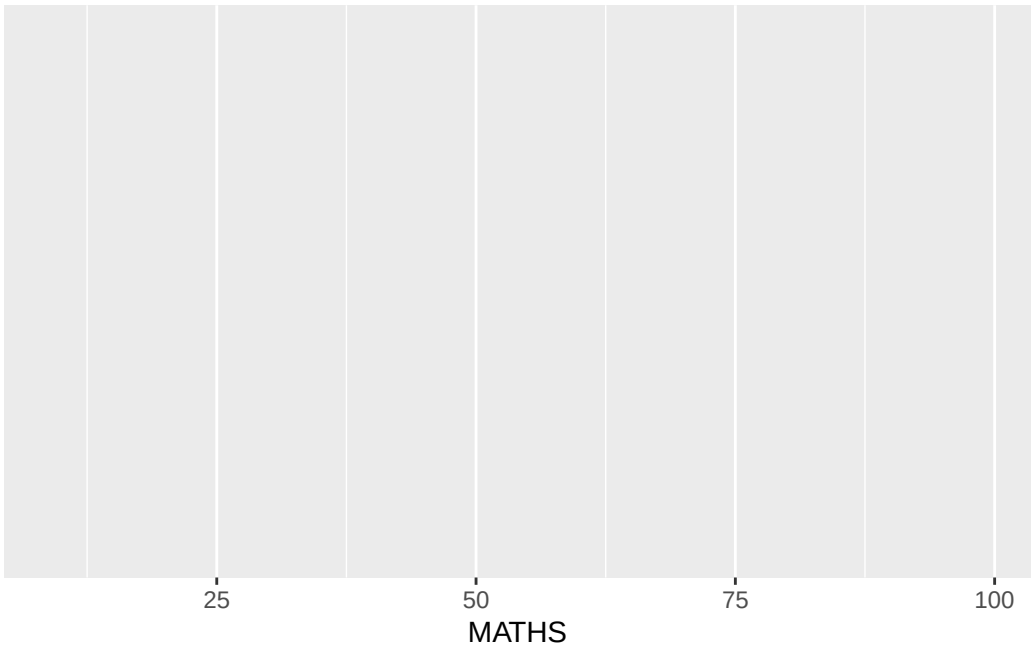


1.5 Essential Grammatical Elements in ggplot2:data

```
ggplot(data=exam_data)
```

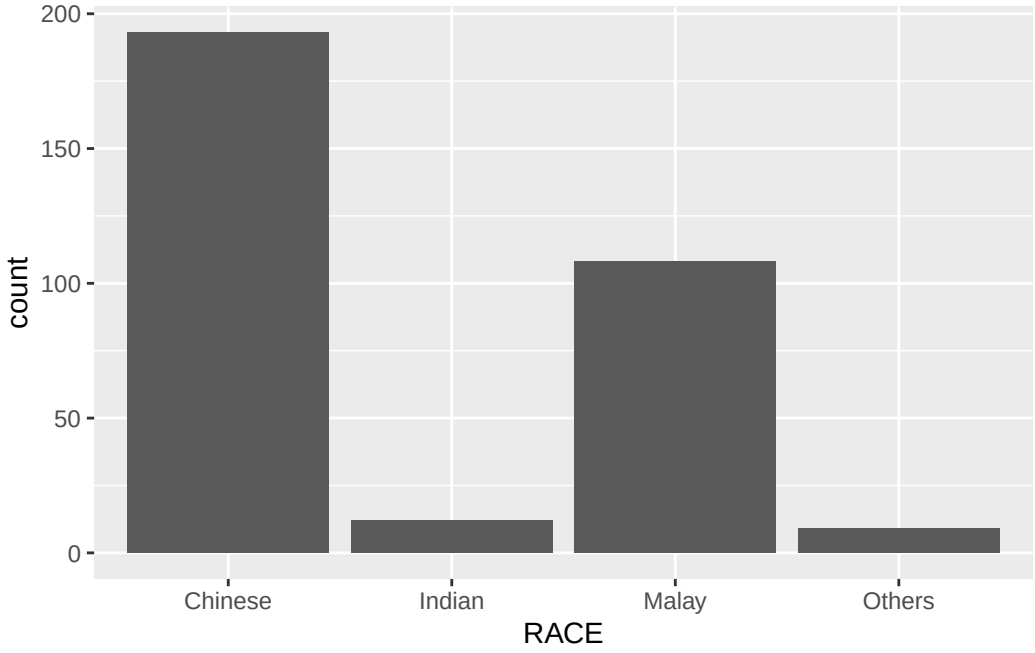
1.6 Essential Grammatical Elements in ggplot2:Aesthetic mappings

```
ggplot(data=exam_data,
       aes(x= MATHS))
```



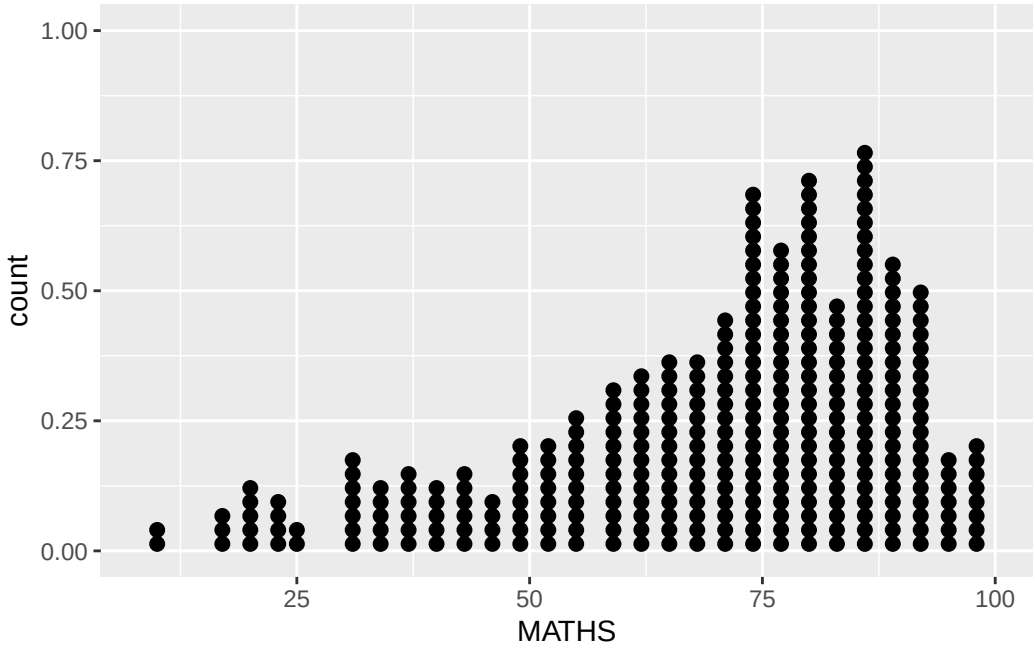
1.7 Essentail Grammatical Elements in ggplot2:geom

```
ggplot(data=exam_data,
       aes(x=RACE)) +
  geom_bar()
```

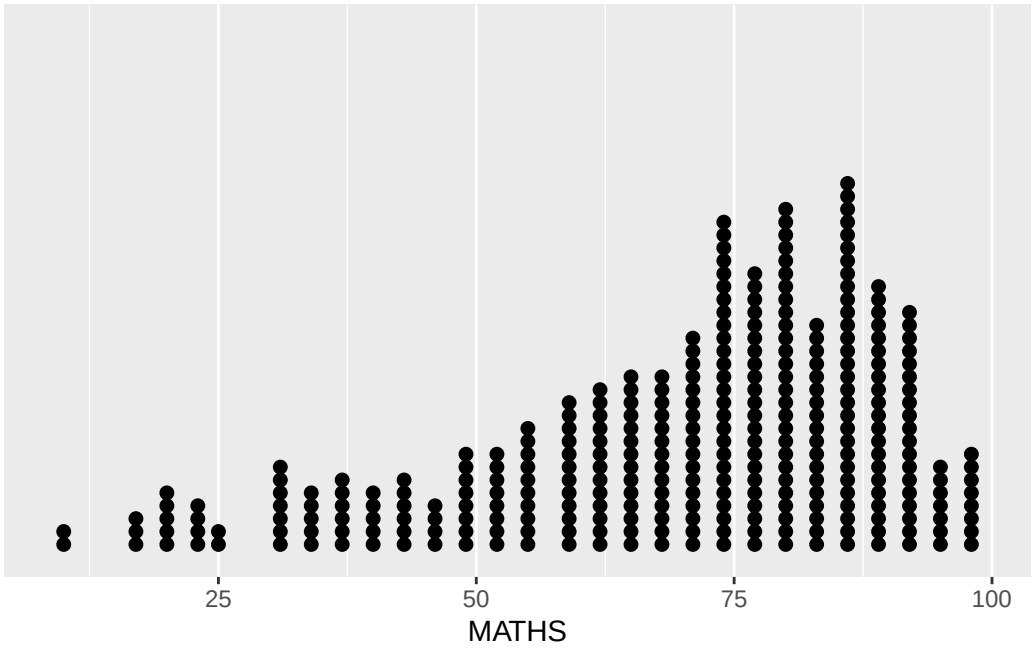



```
ggplot(data=exam_data,
       aes(x = MATHS)) +
  geom_dotplot(dotsize = 0.5)
```

```
Bin width defaults to 1/30 of the range of the data. Pick better value with
`binwidth`.
```



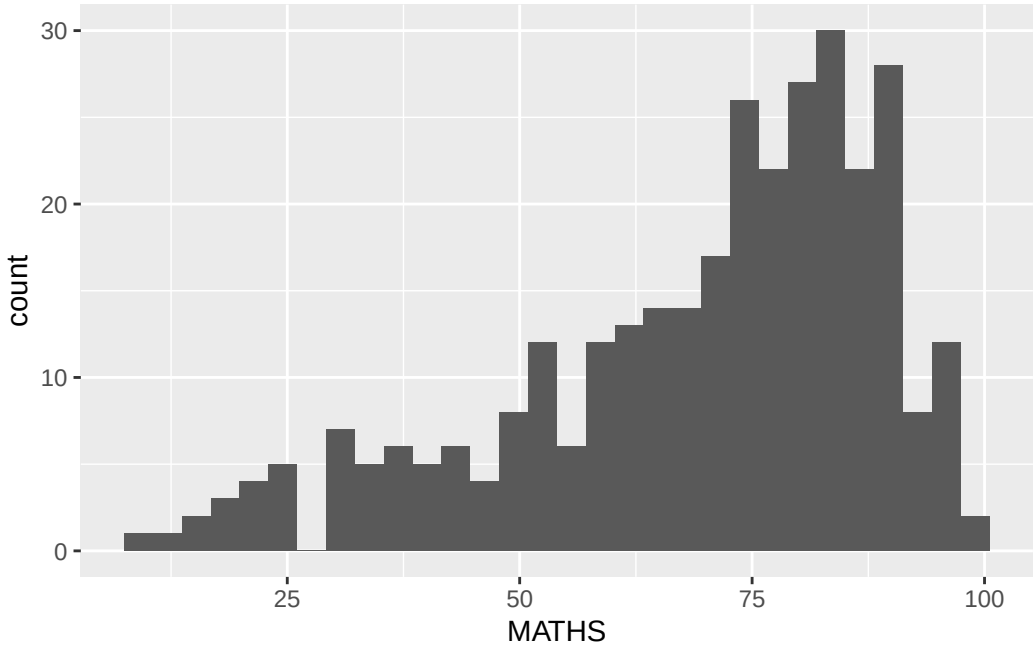

```
ggplot(data=exam_data,
       aes(x = MATHS)) +
  geom_dotplot(binwidth=2.5,
               dotsize = 0.5) +
  scale_y_continuous(NULL,
                     breaks = NULL)
```



1.7.3 Geometric Objects:geom_histogram()

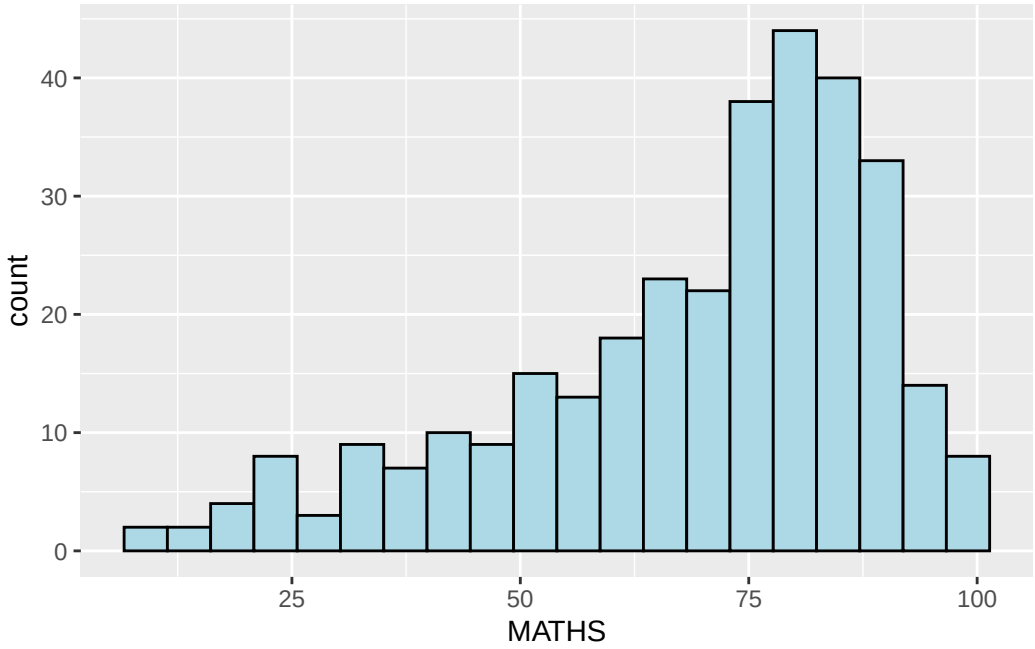
```
ggplot(data=exam_data,
       aes(x = MATHS)) +
  geom_histogram()
```

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



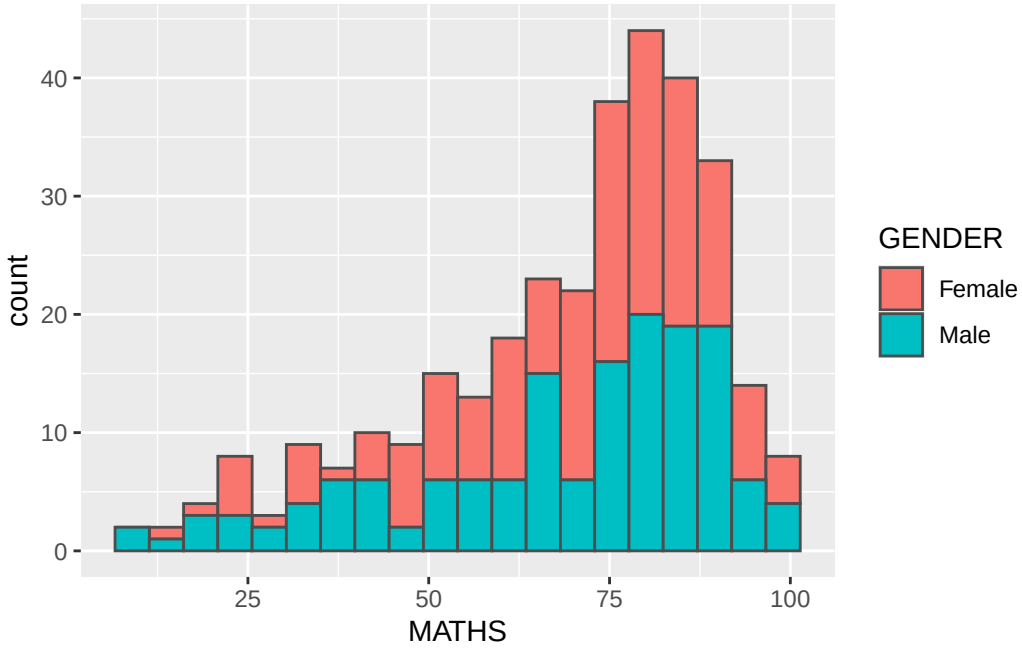
1.7.4 Modifying a geometric object by changing geom()

```
ggplot(data=exam_data,
       aes(x= MATHS)) +
  geom_histogram(bins=20,
                 color="black",
                 fill="light blue")
```



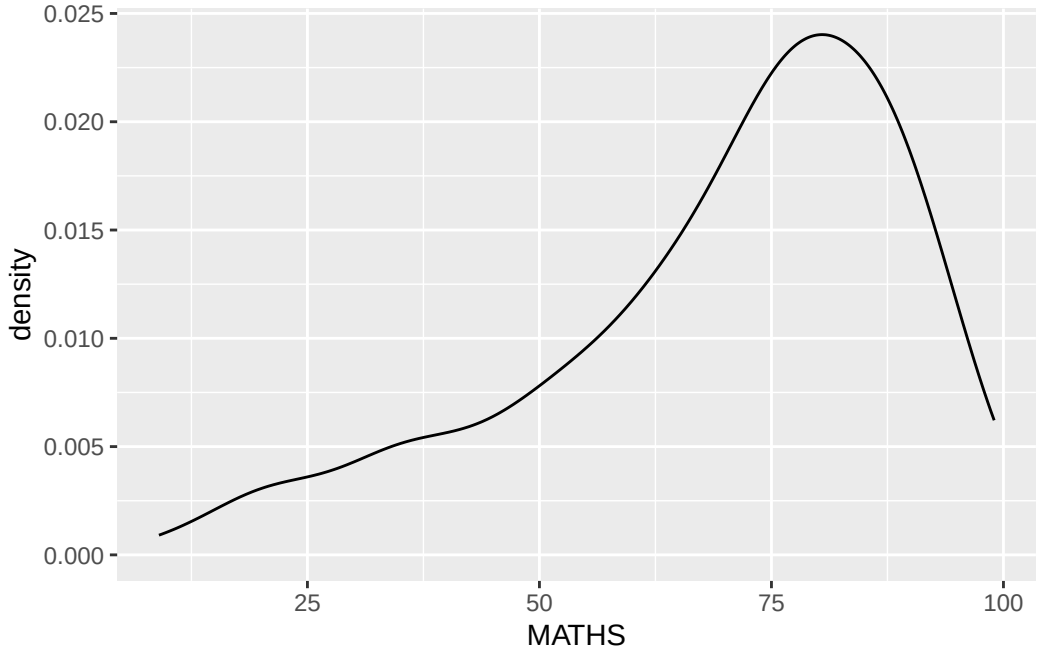
1.7.5 Modifying a geometric object by changing aes()

```
ggplot(data=exam_data,
       aes(x= MATHS,
           fill = GENDER)) +
  geom_histogram(bins=20,
                 color="grey30")
```

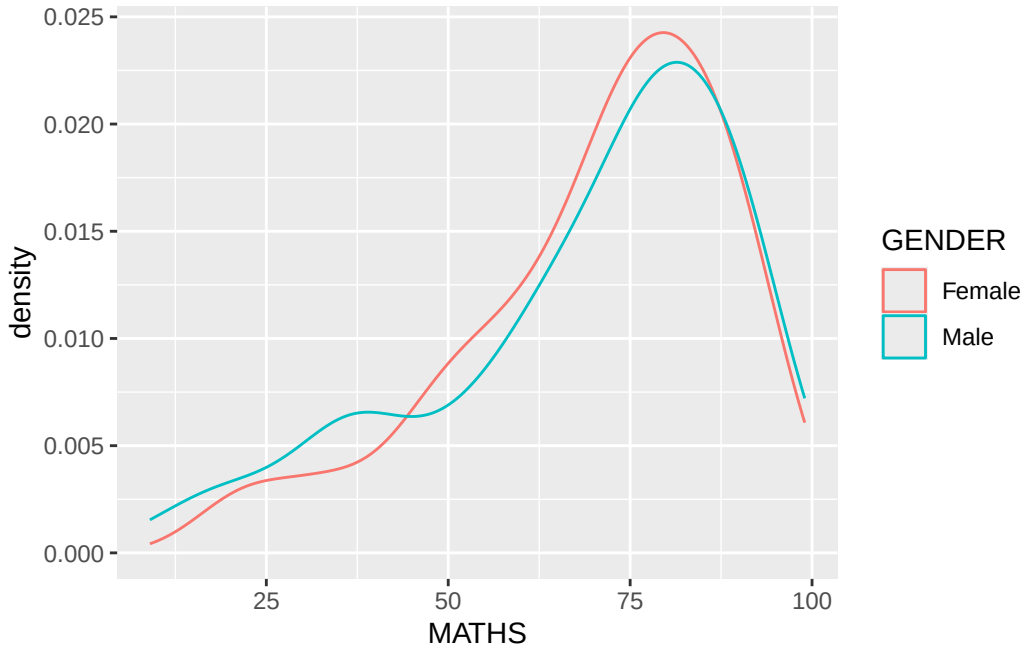


1.7.6 Geometric Objects:geom-density()

```
ggplot(data=exam_data,
       aes(x = MATHS)) +
  geom_density()
```

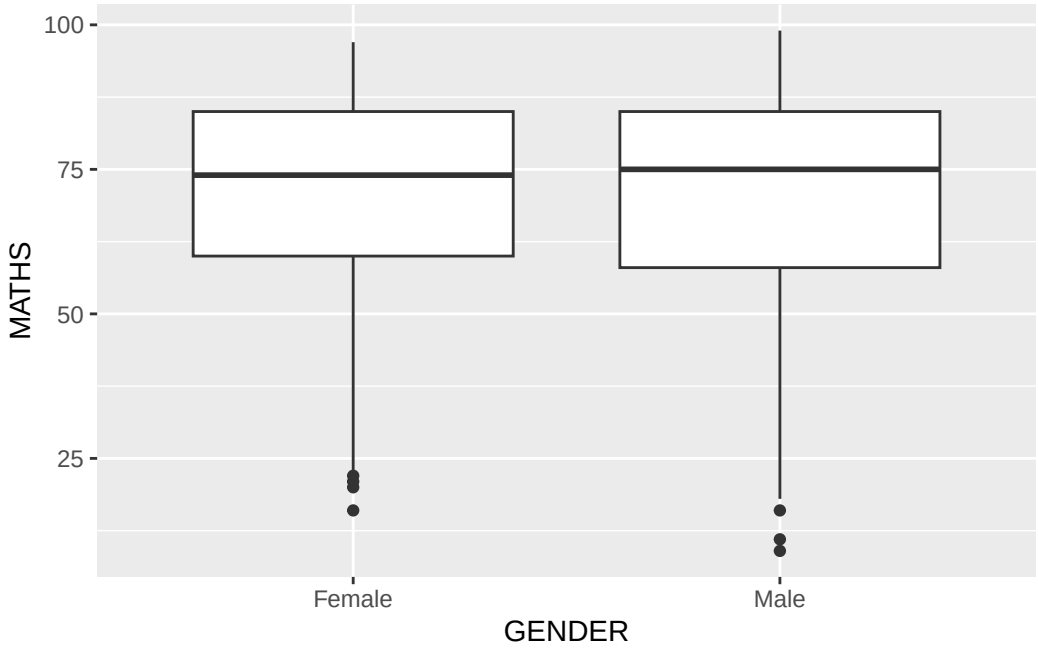



```
ggplot(data=exam_data,
       aes(x = MATHS,
           colour = GENDER)) +
  geom_density()
```

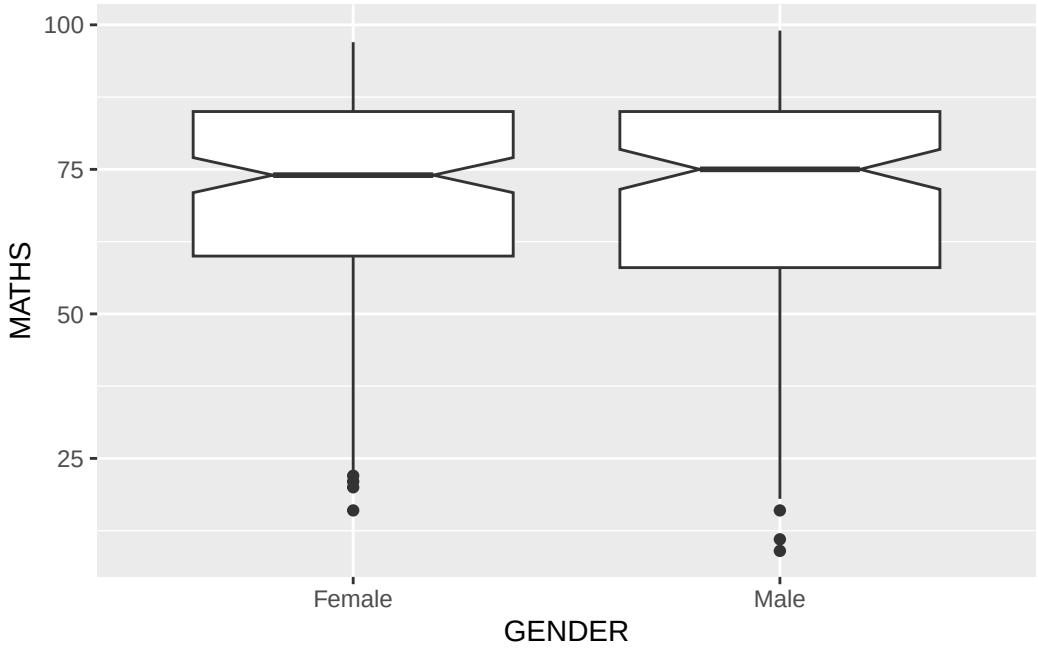


1.7.7 Geometric Objects:geom_boxplot

```
ggplot(data=exam_data,
       aes(y = MATHS,
           x= GENDER)) +
  geom_boxplot()
```

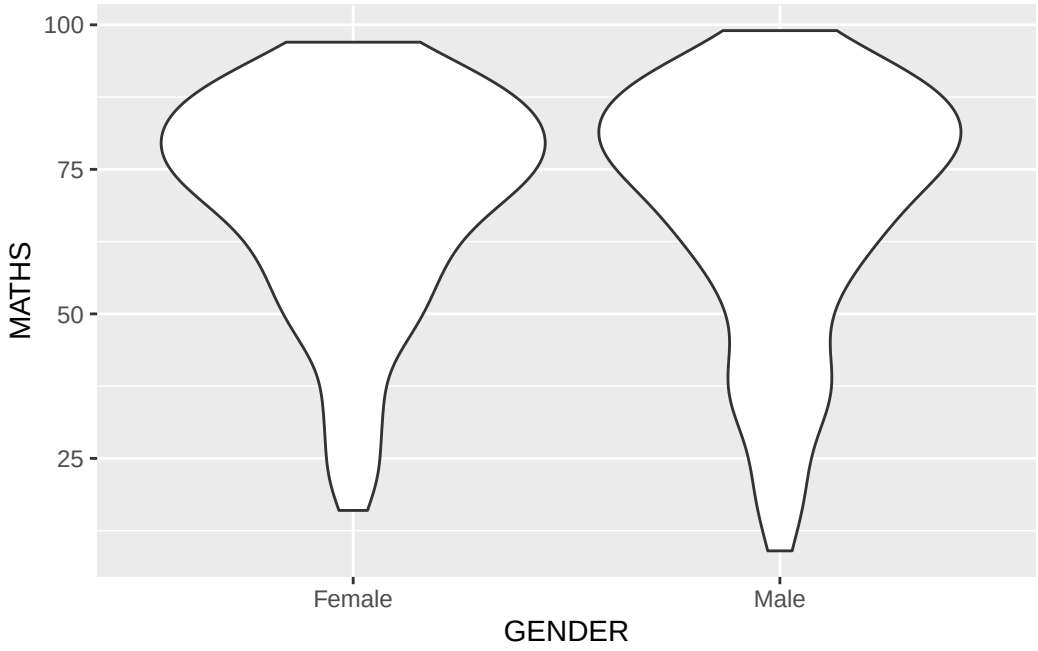



```
ggplot(data=exam_data,
       aes(y = MATHS,
           x= GENDER)) +
  geom_boxplot(notch=TRUE)
```



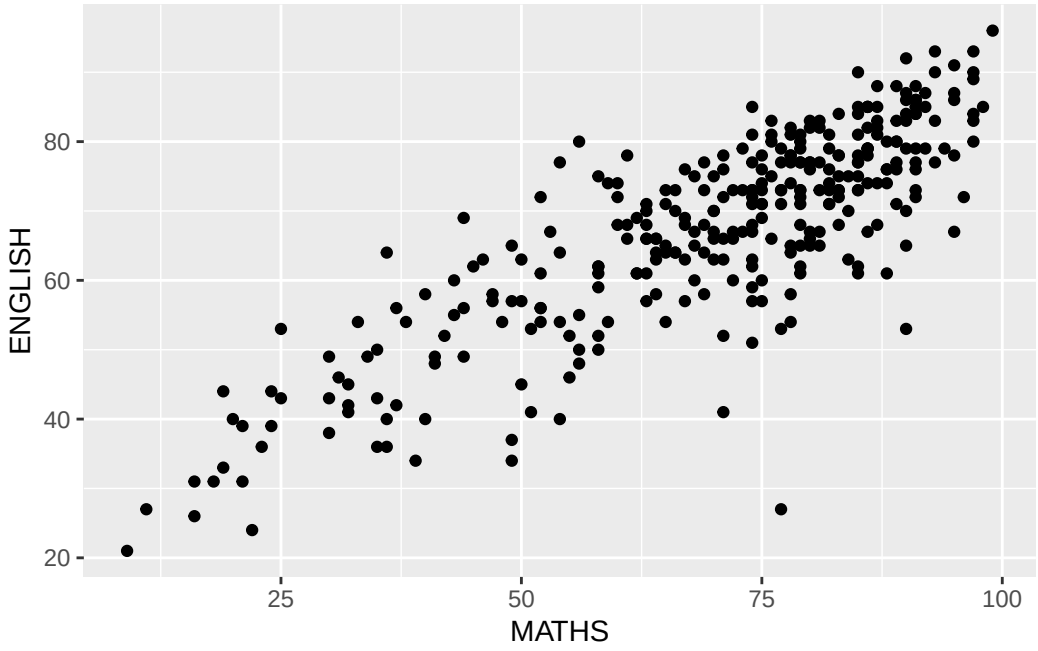
1.7.8 Geometric Objects:geom_violin

```
ggplot(data=exam_data,
       aes(y = MATHS,
           x= GENDER)) +
  geom_violin()
```



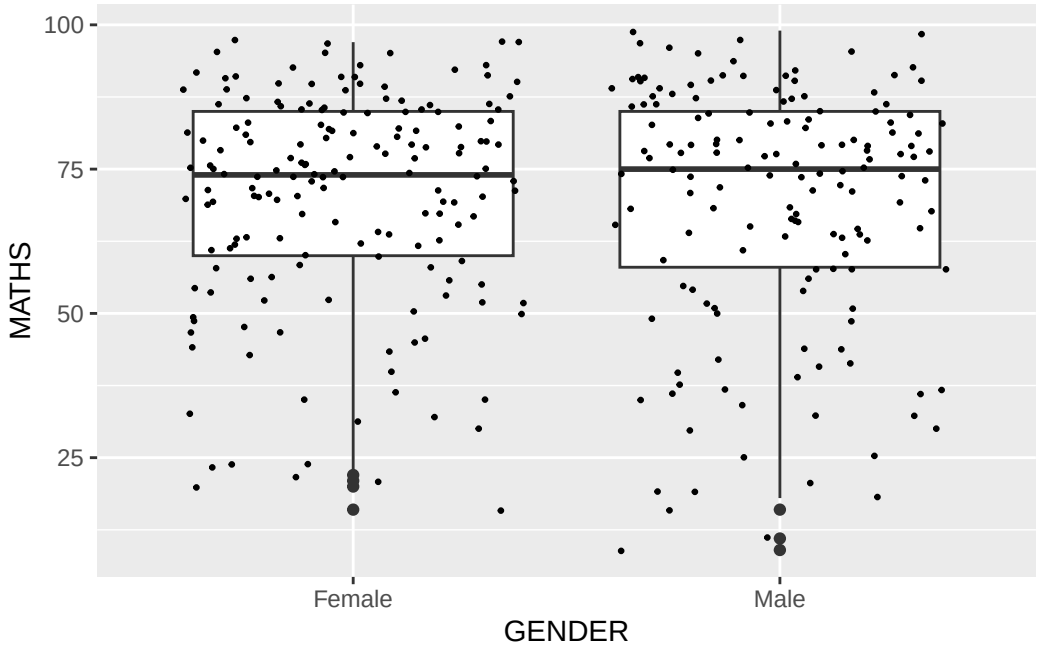
1.7.9 Geometric Objects:geom_point()

```
ggplot(data=exam_data,
       aes(x= MATHS,
           y=ENGLISH)) +
  geom_point()
```



1.7.10 geom objects can be combined

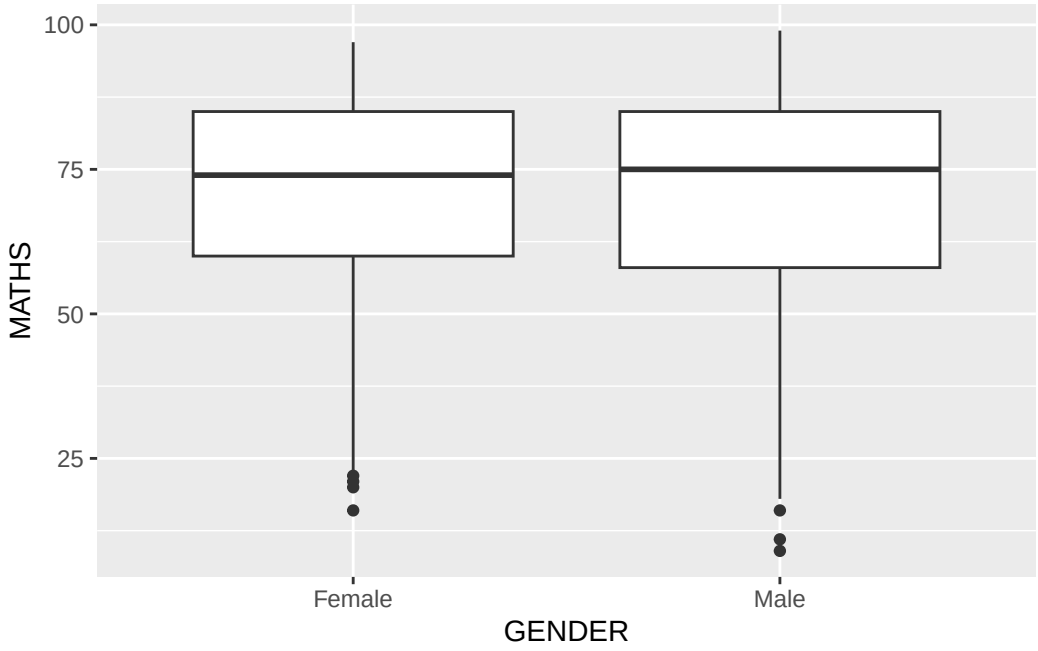
```
ggplot(data=exam_data,
       aes(y = MATHS,
           x= GENDER)) +
  geom_boxplot() +
  geom_point(position="jitter",
             size = 0.5)
```



1.8 Essential Grammatical Elements in ggplot2:stat

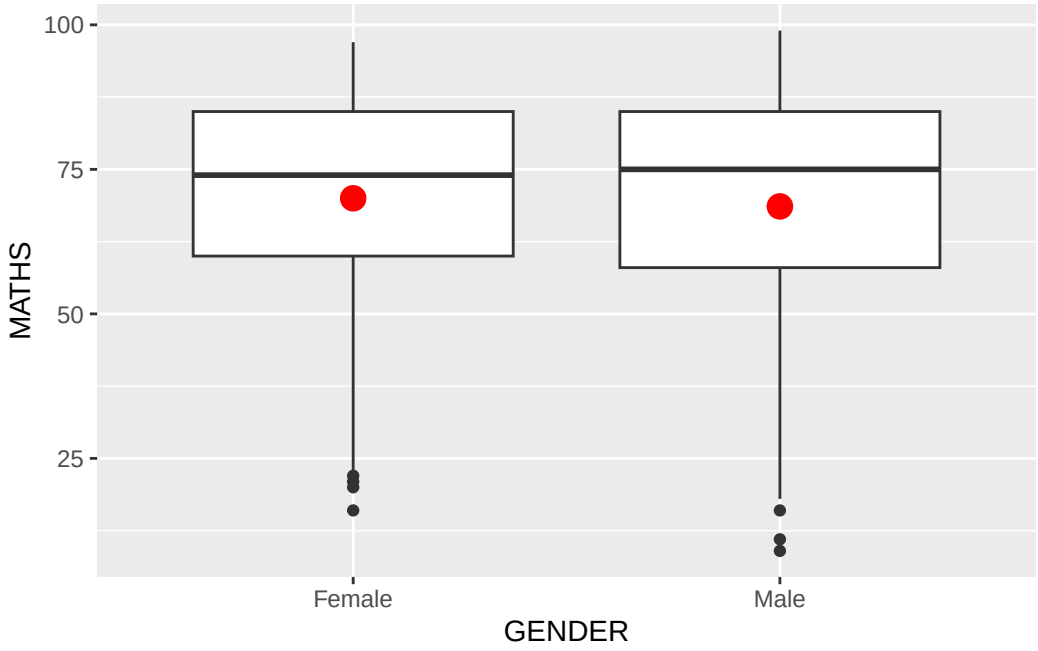
1.8.1 Working with stat()

```
ggplot(data=exam_data,
       aes(y = MATHS, x= GENDER)) +
  geom_boxplot()
```



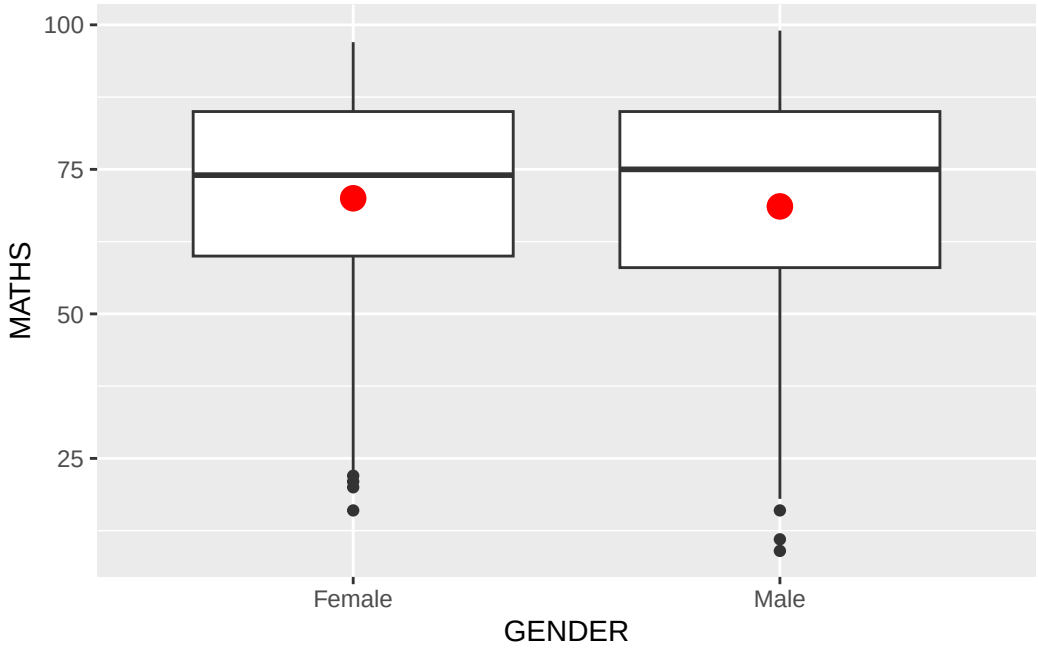
1.8.2 Working with stat()-stat_summary()method

```
ggplot(data=exam_data,
       aes(y = MATHS, x= GENDER)) +
  geom_boxplot() +
  stat_summary(geom = "point",
               fun = "mean",
               colour ="red",
               size=4)
```



1.8.3 Working with stat-the geom()method

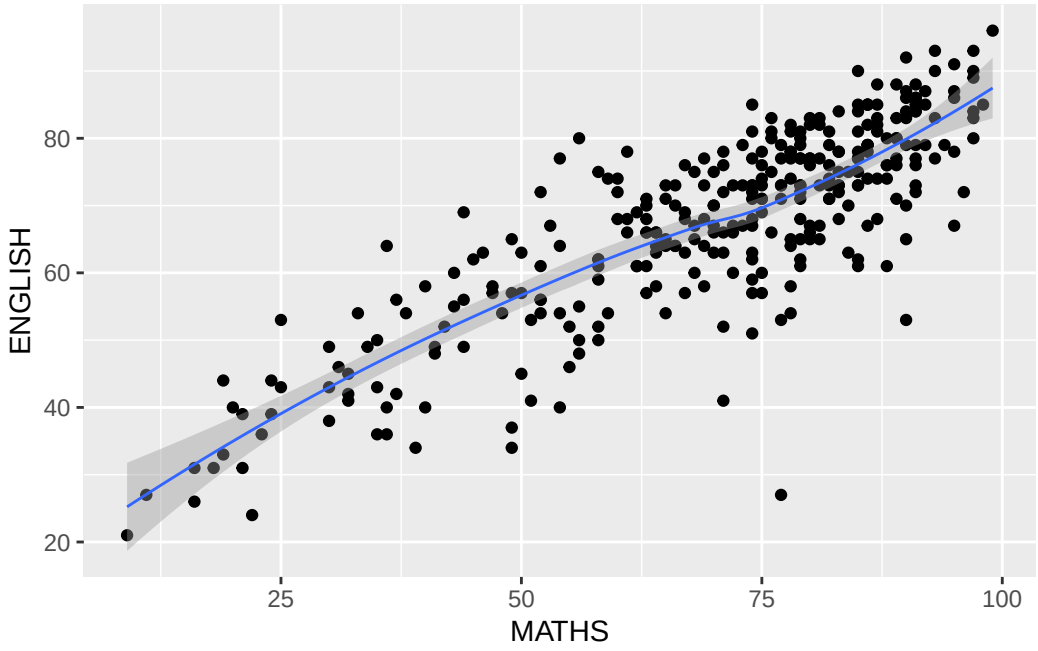
```
ggplot(data=exam_data,
       aes(y = MATHS, x= GENDER)) +
  geom_boxplot() +
  geom_point(stat="summary",
             fun="mean",
             colour="red",
             size=4)
```



1.8.4 Adding a best fit curve on a scatterplot?

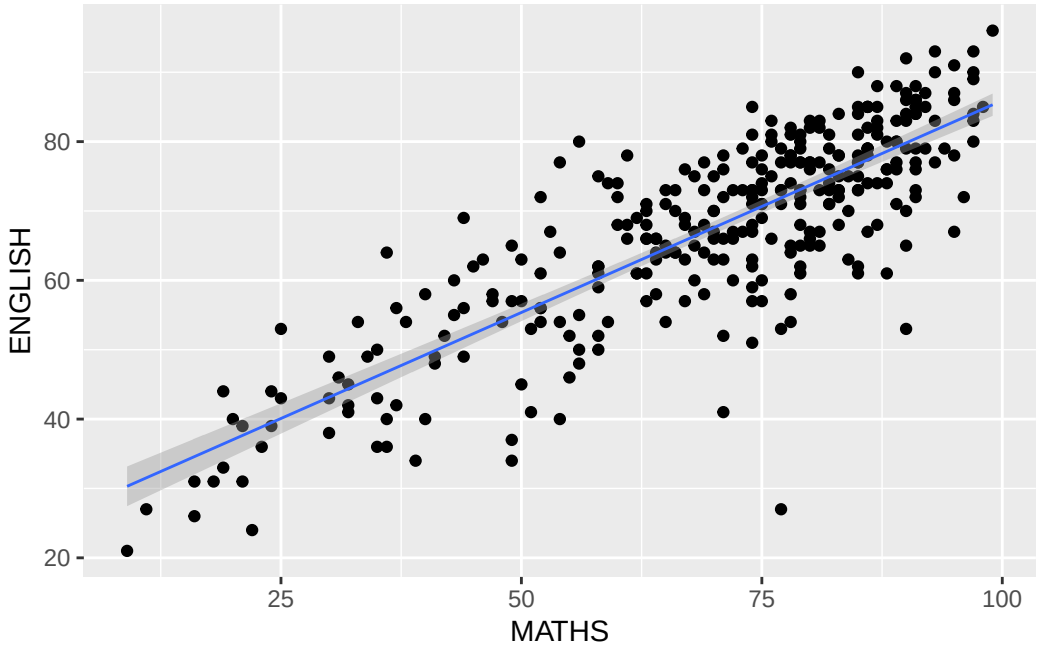
```
ggplot(data=exam_data,
       aes(x= MATHS, y=ENGLISH)) +
  geom_point() +
  geom_smooth(linewidth =0.5)
```

```
`geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```




```
ggplot(data=exam_data,
       aes(x= MATHS,
           y=ENGLISH)) +
  geom_point() +
  geom_smooth(method=lm,
              linewidth=0.5)
```

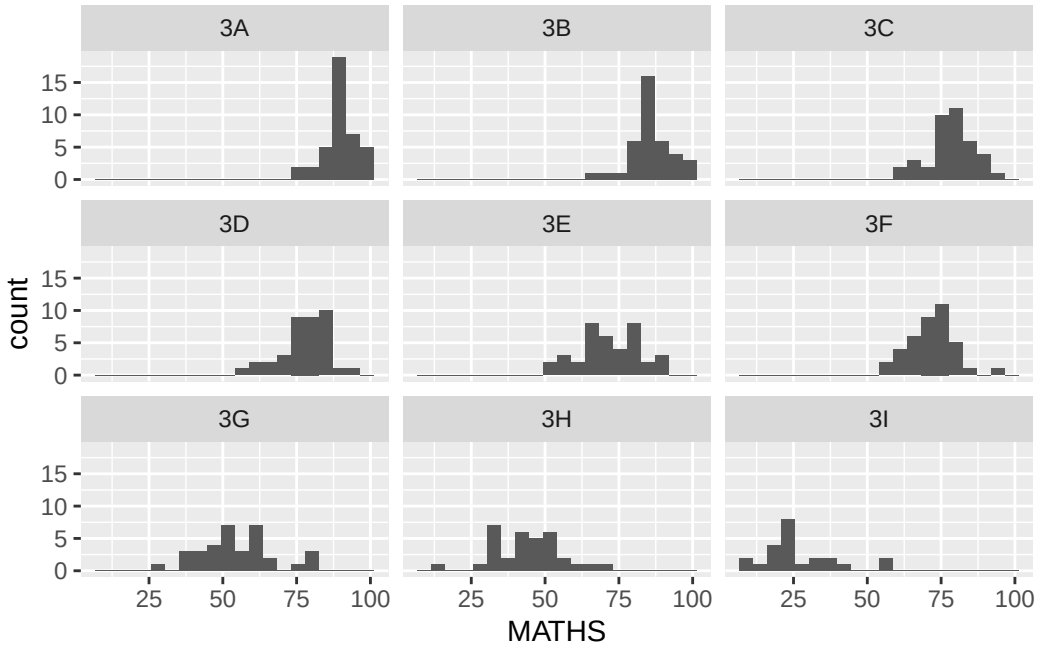
```
`geom_smooth()` using formula = 'y ~ x'
```



1.9 Essential Grammatical Elements in ggplot2:Facets

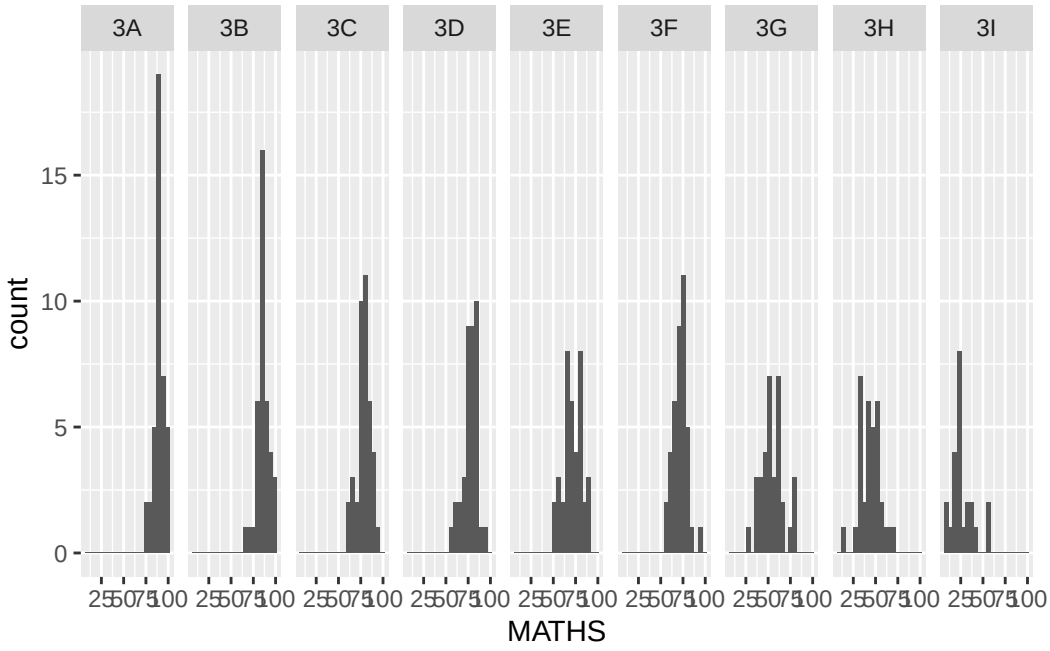
1.9.1 Working with Facet_wrap()

```
ggplot(data=exam_data,
       aes(x= MATHS)) +
  geom_histogram(bins=20) +
    facet_wrap(~ CLASS)
```



1.9.2 facet_grid() function

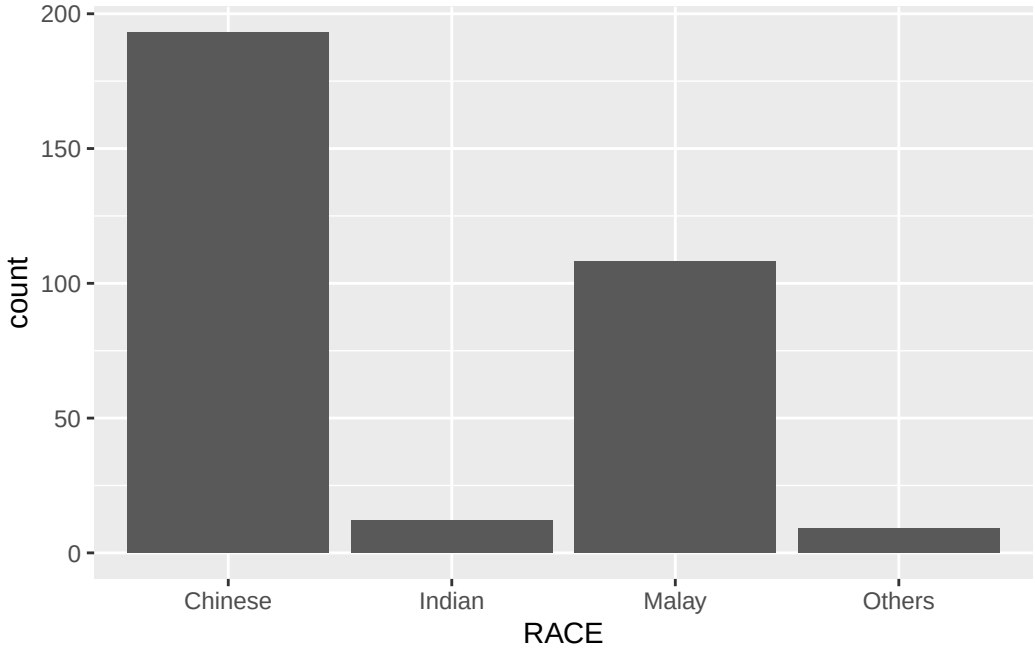
```
ggplot(data=exam_data,
       aes(x= MATHS)) +
  geom_histogram(bins=20) +
    facet_grid(~ CLASS)
```



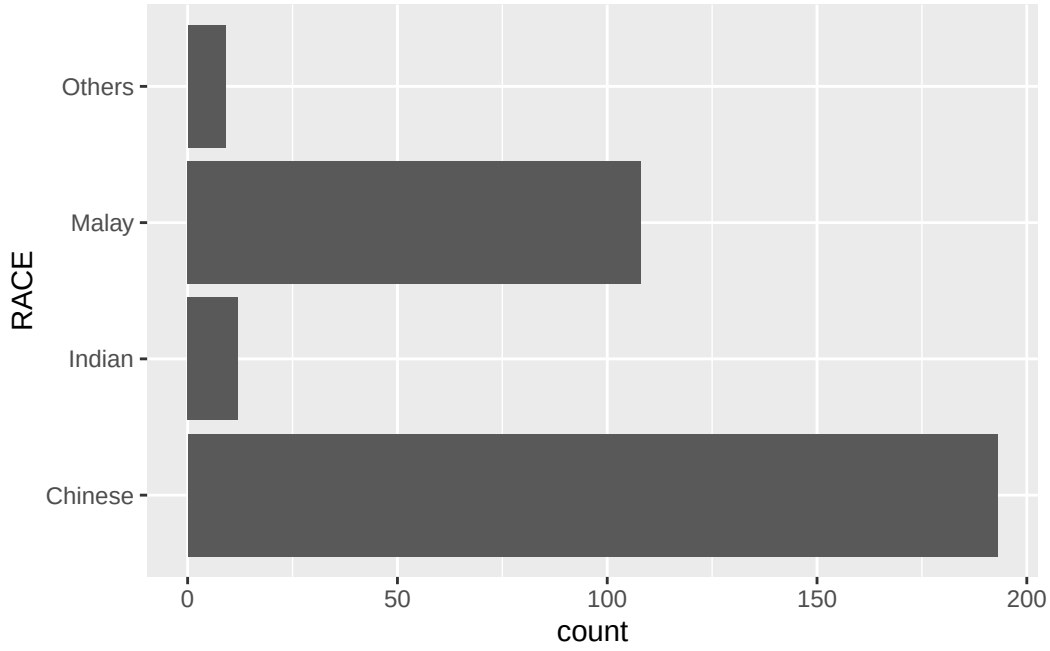
1.10 Essential Grammatical Elements in ggplot2:Coordinates

1.10.1 Working with Coordinate

```
ggplot(data=exam_data, aes(x=RACE)) +
  geom_bar()
```




```
ggplot(data=exam_data,
       aes(x=RACE)) +
  geom_bar() +
  coord_flip()
```

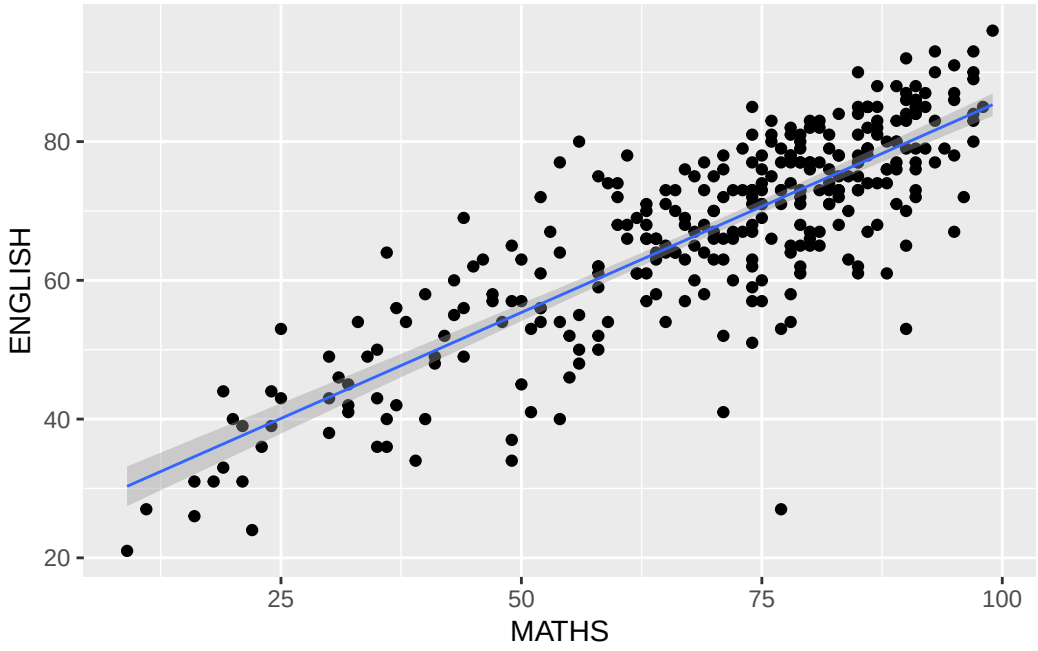


1.10.2 Changing the y-and x-axis range

```
ggplot(data=exam_data,
       aes(x= MATHS, y=ENGLISH)) +
  geom_point() +
  geom_smooth(method=lm, size=0.5)
```

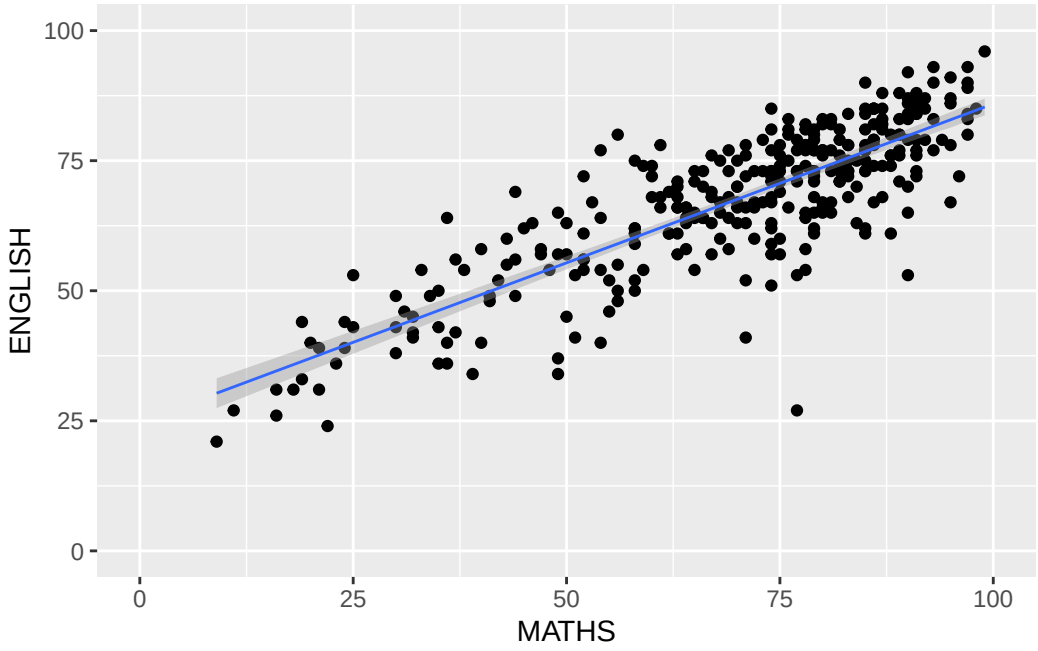
```
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
i Please use `linewidth` instead.
```

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




```
ggplot(data=exam_data,
       aes(x= MATHS, y=ENGLISH)) +
  geom_point() +
  geom_smooth(method=lm,
              size=0.5) +
  coord_cartesian(xlim=c(0,100),
                  ylim=c(0,100))
```

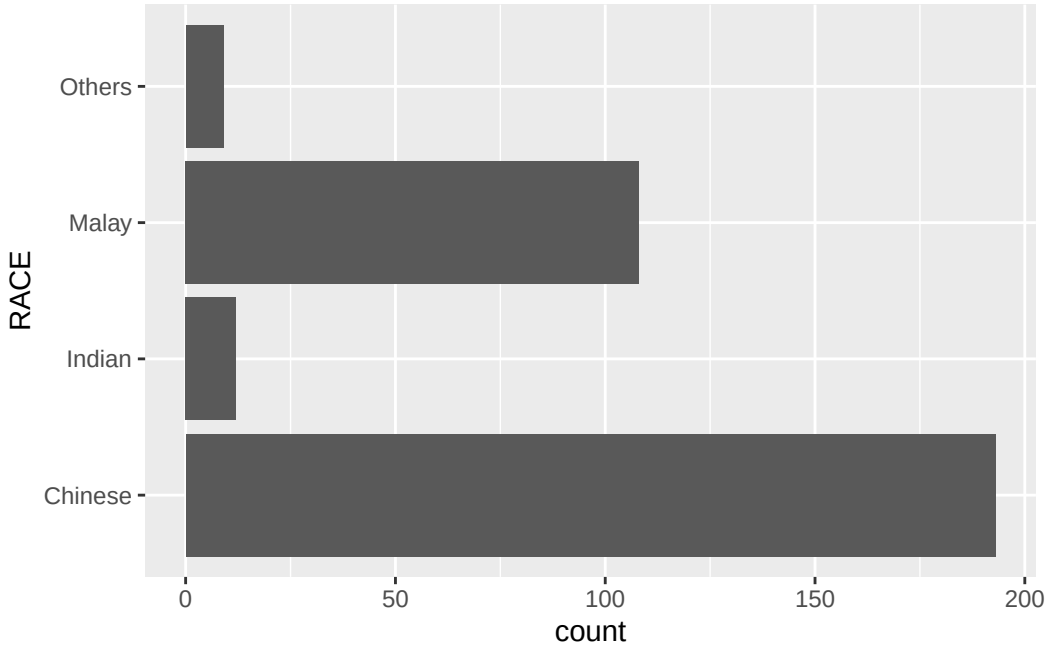
```
`geom_smooth()` using formula = 'y ~ x'
```



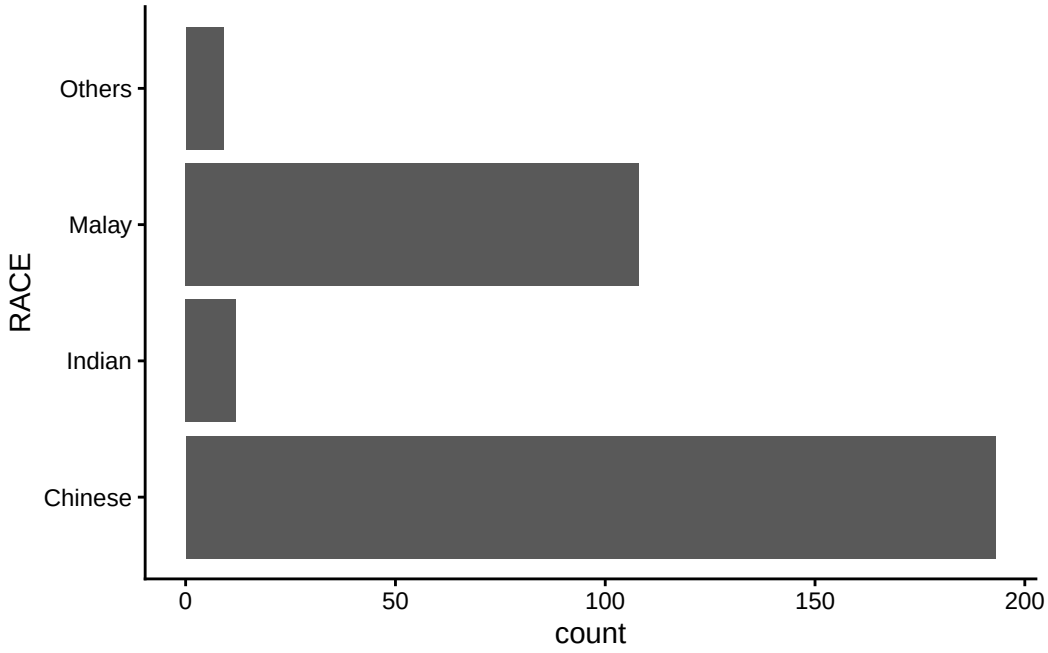
1.11 Essential Grammatical Elements in ggplot2:themes

1.11.1 Working with theme

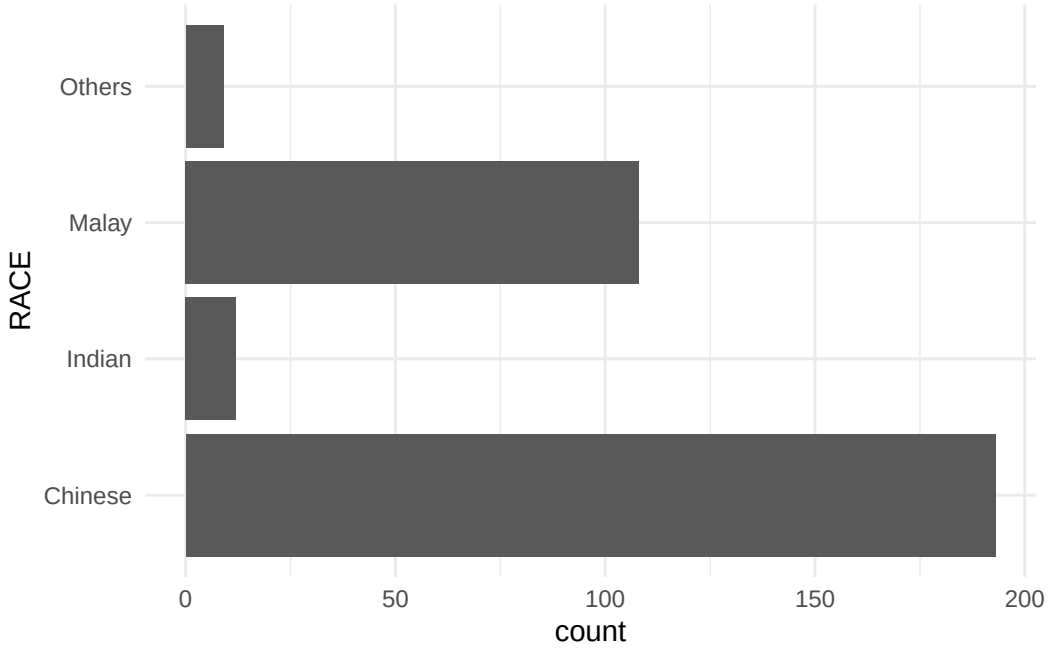
```
ggplot(data=exam_data,
       aes(x=RACE)) +
  geom_bar() +
  coord_flip() +
  theme_gray()
```




```
ggplot(data=exam_data,
       aes(x=RACE)) +
  geom_bar() +
  coord_flip() +
  theme_classic()
```




```
ggplot(data=exam_data,
       aes(x=RACE)) +
  geom_bar() +
  coord_flip() +
  theme_minimal()
```


Others
Malay
Indian
Chinese
RACE
0
50
100
150
200
count