Data Visualization

Byteflow Dynamics

9/24/2017

Tidyverse

Tidyverse is a collection of R package for data science. It includes

- ggplot2
- tidyr
- dplyr

among other packages.

To install tidyverse (or any other package)

- click packages on the right side of Rstudio
- click install, type 'tidyverse', click install

To load tidyverse (or any other package)

```
library(tidyverse)
```

```
## Loading tidyverse: ggplot2
## Loading tidyverse: tibble
## Loading tidyverse: tidyr
## Loading tidyverse: readr
## Loading tidyverse: purrr
## Loading tidyverse: dplyr
```

```
## Warning: package 'tidyr' was built under R version 3.4.1
```

```
## Warning: package 'purrr' was built under R version 3.4.1
```

```
## Conflicts with tidy packages ----------------------------------------------
```

```
## filter(): dplyr, stats
## lag():    dplyr, stats
```

ggplot

ggplot is one of the most versatile and widely used graphing system in R.

ggplot template

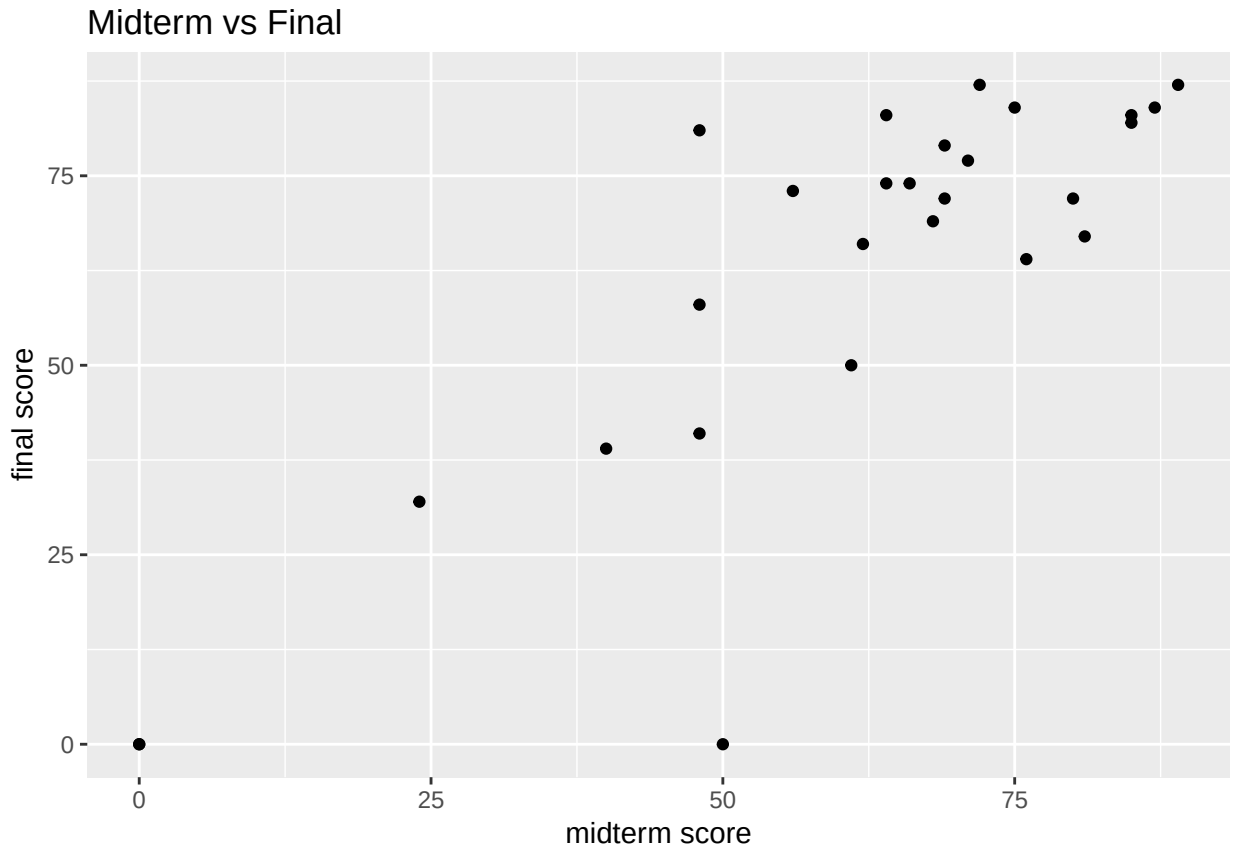
```
ggplot(data = <DATA>) +
  <GEOM_FUNCTION>(mapping = aes(<MAPPINGS>))
```

To make a graph, replace the bracketed sections in the code above with a dataset, a geom function, or a collection of mappings.

Scatter plot

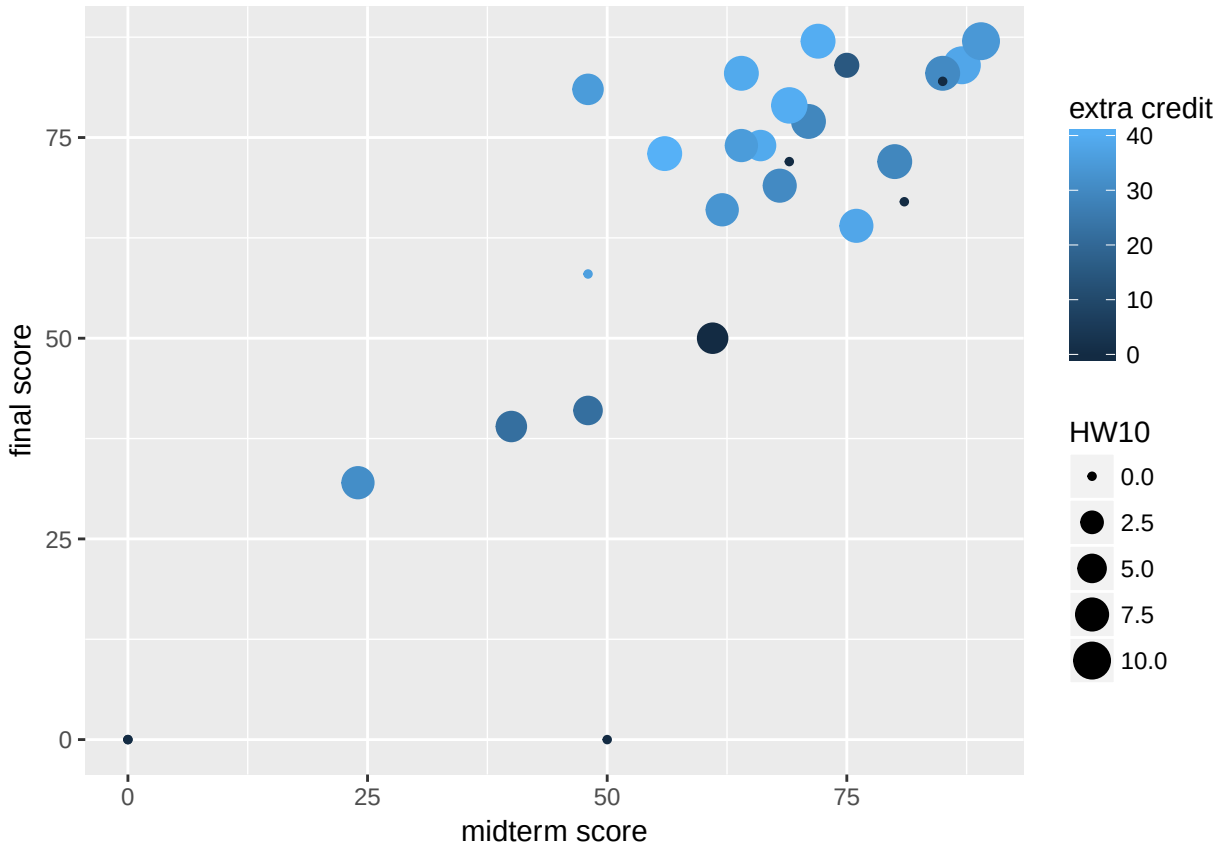
Example: Find if there's a correlation between midterm and final grades by making a scatter plot.

```
ggplot(data = lec) +
  geom_point(mapping = aes(x = lec$Midterm, y = lec$Final.exam)) +
  labs(x = "midterm score", y ="final score") +
  ggtitle("Midterm vs Final")
```



If you want to add another dimension, you can add color or size.

```
ggplot(data = lec) +
  geom_point(mapping = aes(x = Midterm, y = lec$Final.exam, color = lec$Extra.Credit, size = lec$HW.Ch.
  labs(x = "midterm score", y = "final score", color = "extra credit", size = "HW10")
```



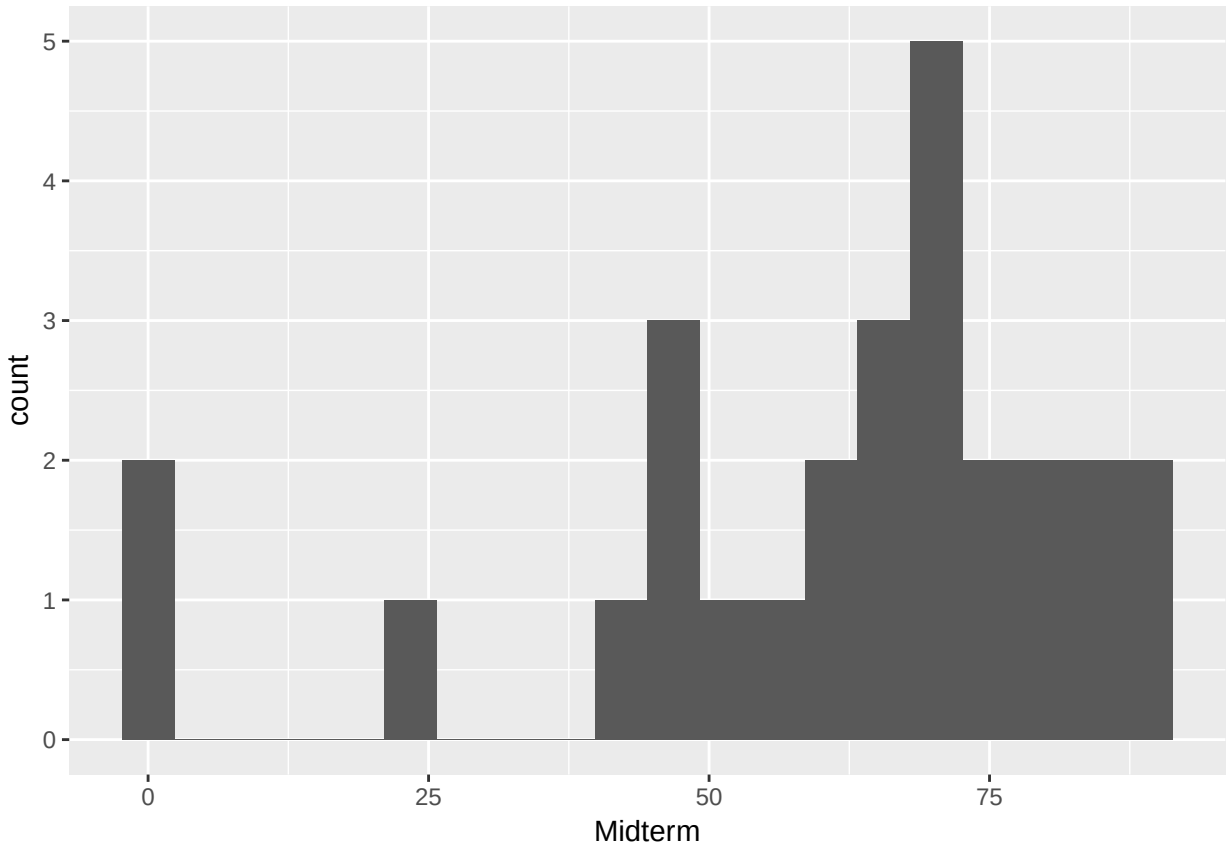
Exercise: Make a scatter plot using the lecture data to find if there's any correlation between the two quiz scores (Quiz Ch 2 and Ch 3)

Extra: Add HW Ch1 score as the color variable

Histogram

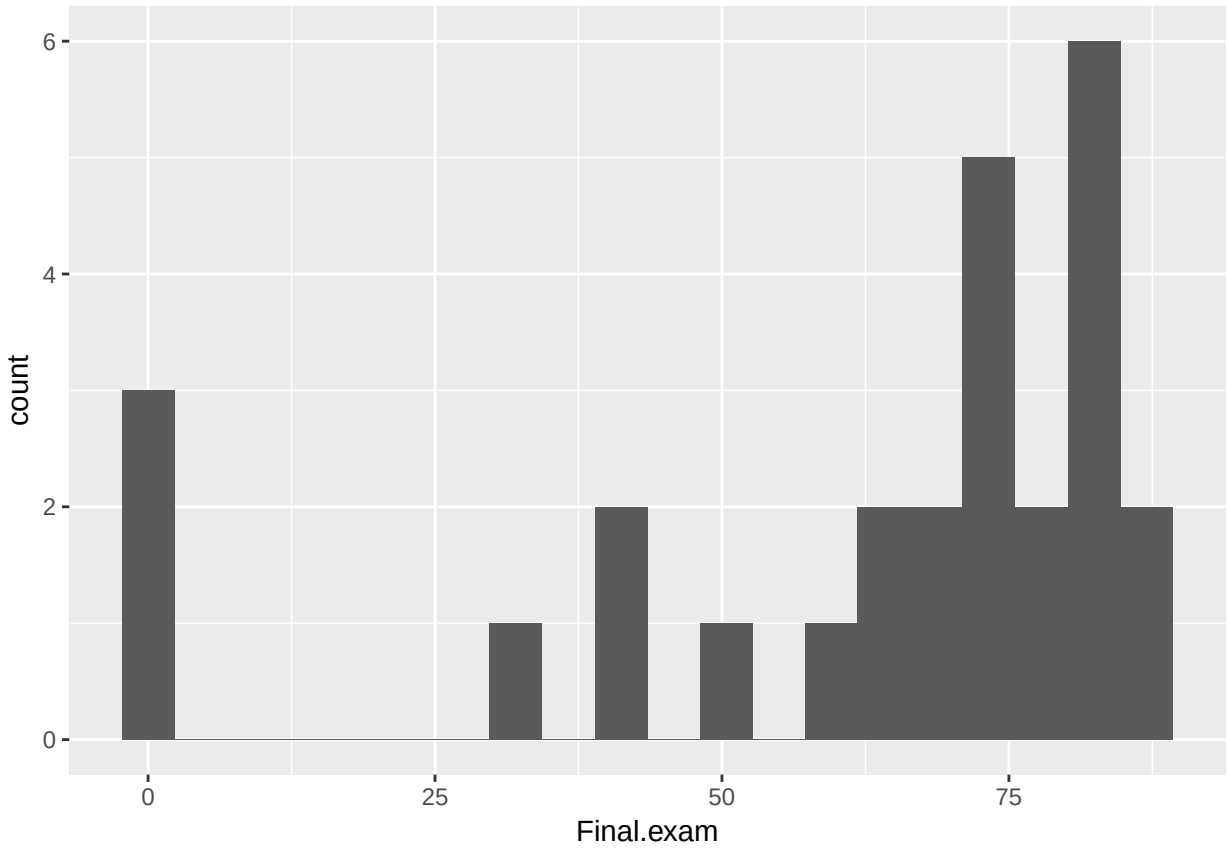
Example: Using the lecture data, make a histogram of students' midterm scores.

```
ggplot(data = lec) +
  geom_histogram(mapping = aes(x = Midterm), bins = 20)
```



Exercise: Using the lecture data, make a histogram of final exam scores.

```
ggplot(data = lec) +
  geom_histogram(mapping = aes(x = Final.exam), bins =20)
```


6
4
2
0
count
0
25
50
75
Final.exam