R Cheatsheet for EDUC 6600

Preparing Your Data

```
**Before starting, load the tidyverse: library(tidyverse)**
```

1

Reading in Your Data

If your data is a **CSV**

```
library(rio)
d <- import("path/to/my_file.csv")
```

If your data is an **SPSS** file

```
library(rio)
d <- import("path/to/my_file.sav")
```

If your data is an **Excel** file

```
library(rio)
d <- import("path/to/my_file.xlsx")
```

If your data is in other formats, come talk to me or Google it

2.1

Clean Your Data (Part 1)

The **pipe** %>% takes what is before it and brings it into the next function (can be read as "then")

```
# Not real code
me %>% wake_up() %>% exercise() %>%
  eat_breakfast() %>% shower()
```

Mutate means to *add* or *change* a variable

```
d_new <- d %>%
  mutate(salary_yr = salary_mnth * 12)
```

Filter means to *subset the rows*

```
d_females <- d %>%
  filter(sex == "female")
```

2.2

Clean Your Data (Part 2)

factor() creates a categorical variable (nominal)

```
d_new <- d %>%
  mutate(var_f = factor(var))
```

case_when() gives values based on a condition

```
d_new <- d %>%
  mutate(v2 = case_when(var == 1 ~ "val"))
```

To get the mean of *each observation* across vars

```
library(furniture)
d_new <- d %>%
  mutate(mean = rowmeans(var1, var2))
```

3

Check Your Data

Get **frequencies** of your variables

```
library(furniture)
tableF(data, var1)
```

Major	Freq	CumFreq	Percent	CumPerc
Psychology	29	29	29.00%	29.00%
Premed	25	54	25.00%	54.00%
Biology	21	75	21.00%	75.00%
Sociology	15	90	15.00%	90.00%
Economics	10	100	10.00%	100.00%

Get **means and SDs** of many (or all) of your variables

```
library(furniture)
table1(data,
       var1, var2, var3)
```

See your data

```
View(data)
```

Get a bunch of info

```
library(psych)
describe(data)
```