



# Welcome!

---

Welcome to *week 8*!

**Record the meeting**

# Breakout rooms!

---

Starting with whomever has the most bandwidth at home/outside of school (and work):

## **One question:**

- What is something you're grateful for?

## **One reflection/discussion:**

- What is one thing that you like about this course? And, what is one thing about this course that you wished were different?

# Review of last week's class

- Using color
- Grouping and stacking bar charts
- Faceting charts

# Sometimes, a table says it better

From webinar: <https://jrosen48.github.io/airs-ngss-survey/>

Recording: <https://www.youtube.com/watch?v=WxdWzTlzYml>

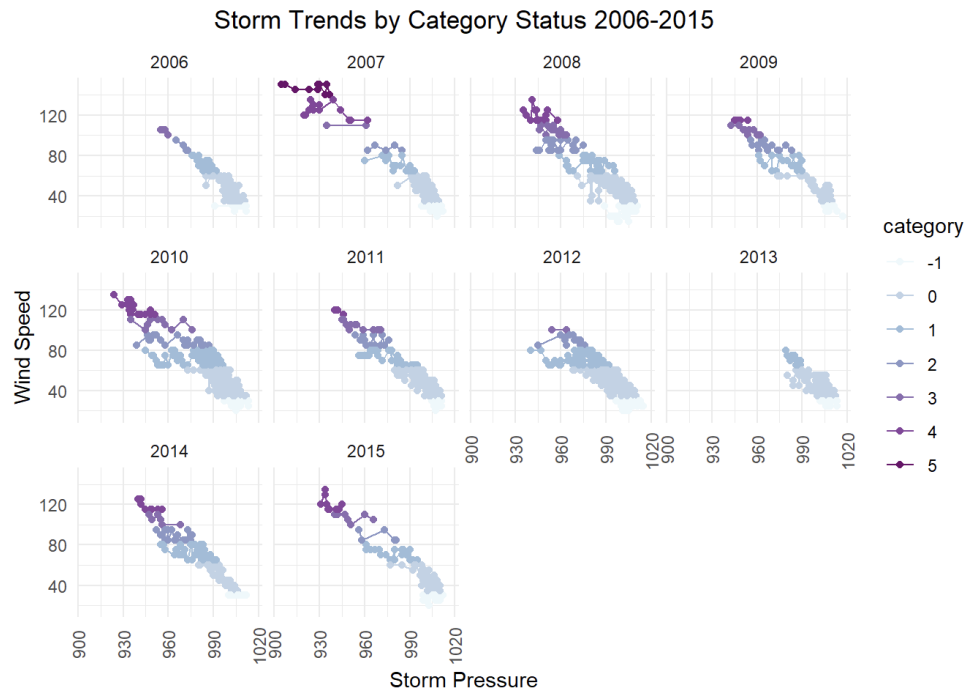
**Table 1**

*Percentage of Teachers Reporting Their School is Teaching the NGSS by State by NGSS Adoption Status*

<b>Adoption Status</b>	<b># of States in AIRS Sample</b>	<b>Mean % (2019)</b>	<b>Mean % (2020)</b>	<b>Mean % (Both Years)</b>
The NGSS	5	78%	86%	82%
Standards Based on the NGSS <input type="checkbox"/>	6	49%	57%	53%
Neither Based on the NGSS nor the NGSS	1	64%	61%	63%
<i>Total (Mean %)</i>	12	61%	67%	64%

# Homework highlights

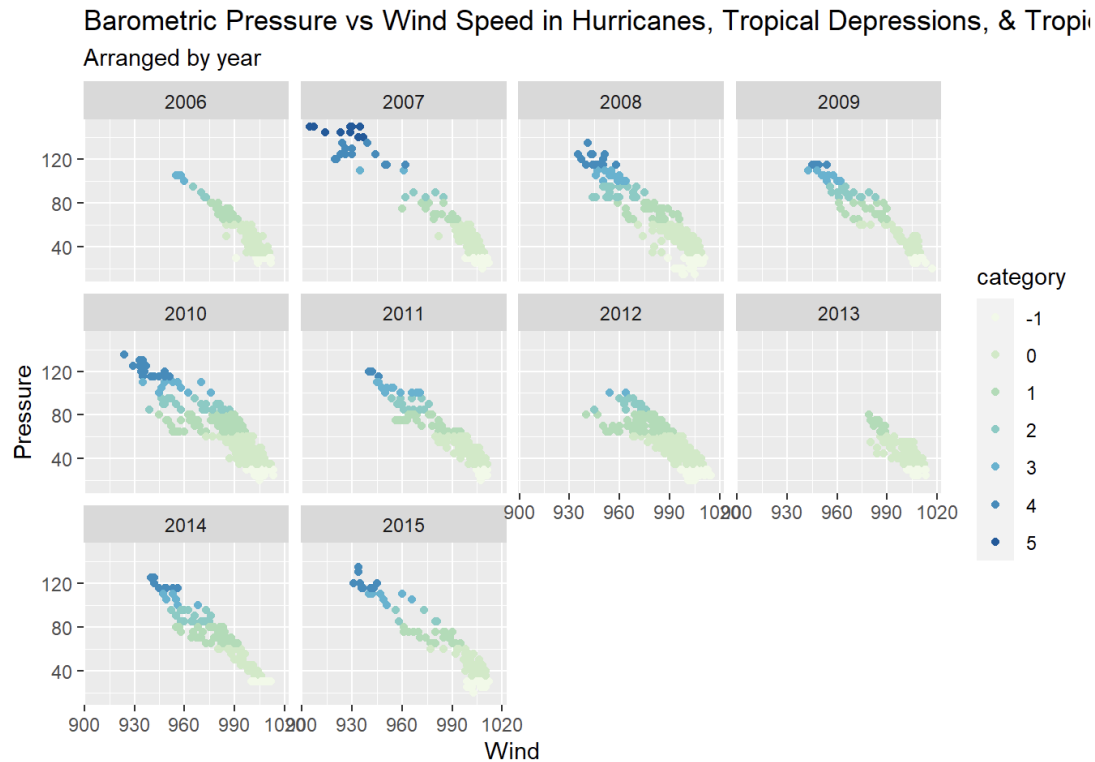
```
FilteredStorms %>%  
  ggplot(aes(x= pressure, y= wind, color= category)) +  
  geom_point()+  
  geom_line()+  
  theme_minimal()+  
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 0.7))+  
  scale_color_brewer(type = "seq", palette = "BuPu")+  
  facet_wrap(~year)+  
  xlab("Storm Pressure")+  
  ylab("Wind Speed")+  
  ggtitle("Storm Trends by Category Status 2006-2015")+  
  ggeasy::easy_center_title()
```



```

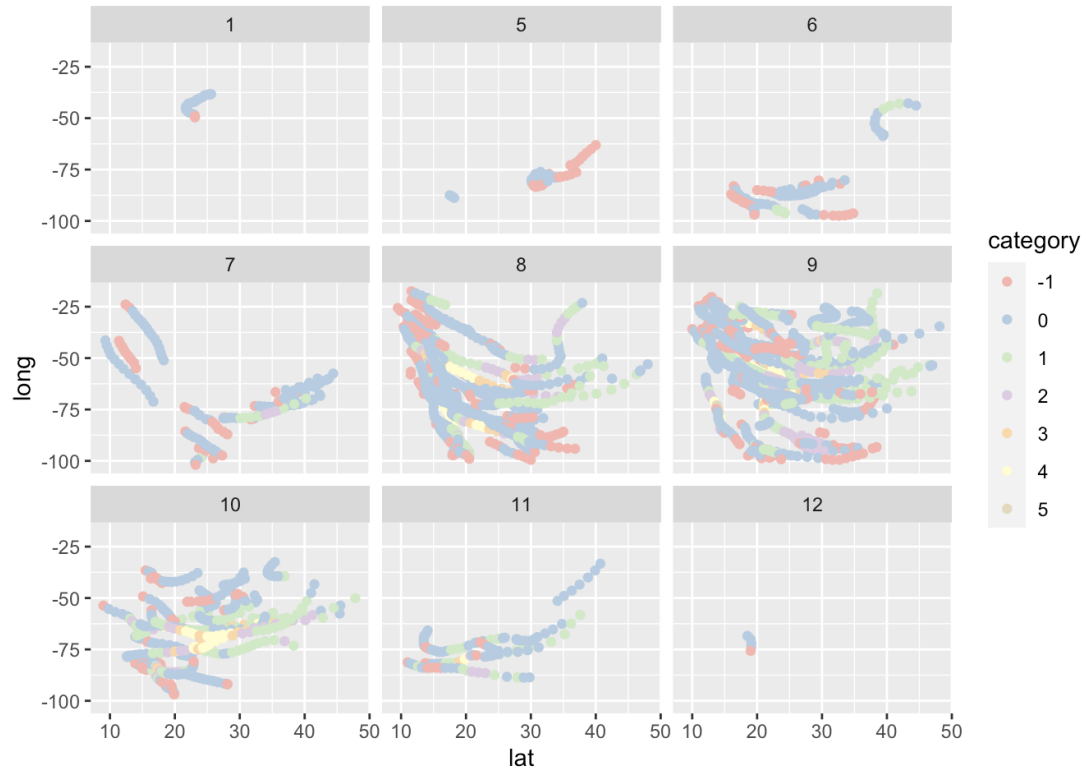
storms_2006 %>%
  ggplot(aes(pressure, wind, color = category)) +
  geom_point() +
  scale_color_brewer(type = "seq", palette = "GnBu") +
  facet_wrap(~year)+
  xlab("Wind") +
  ylab("Pressure") +
  ggtitle("Barometric Pressure vs Wind Speed in Hurricanes, Tropical Depressions, & Tropical Storms by Category")
+
  labs(subtitle = "Arranged by year")

```





```
ggplot(St,aes(x=lat,y=long,color=category))+
  geom_point()+
  scale_color_brewer(type="qual",palette = 4)+
  facet_wrap(~month)
```



# This week's topics

## **Overview**

- A. Summarizing variables
- B. Inspecting data frames and correlations
- C. Summary tables

# A. Summarizing variables

## Overview

- Sometimes you need to summarize variables
- Data viz is often a good tool to use
- Statistics are important too

As always with R, there are multiple ways to get there.

# A. Summarizing variables – Base R

library()

```
summary(mtcars)
```

```
##           mpg           cyl           disp           hp
##  Min.      :10.40   Min.       :4.000   Min.       : 71.1   Min.       : 52.0
## 1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5
## Median :19.20   Median :6.000   Median :196.3   Median :123.0
## Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7
## 3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0
## Max.    :33.90   Max.    :8.000   Max.    :472.0   Max.    :335.0
##           drat           wt           qsec           vs
##  Min.      :2.760   Min.      :1.513   Min.      :14.50   Min.      :0.0000
## 1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000
## Median :3.695   Median :3.325   Median :17.71   Median :0.0000
## Mean   :3.597   Mean   :3.217   Mean   :17.85   Mean   :0.4375
## 3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000
## Max.    :4.930   Max.    :5.424   Max.    :22.90   Max.    :1.0000
##           am           gear           carb
##  Min.      :0.0000   Min.      :3.000   Min.      :1.000
## 1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000
## Median :0.0000   Median :4.000   Median :2.000
## Mean   :0.4062   Mean   :3.688   Mean   :2.812
## 3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000
## Max.    :1.0000   Max.    :5.000   Max.    :8.000
```

# A. Summarizing variables – skimr

skimr package – <https://cran.r-project.org/web/packages/skimr/vignettes/skimr.html>












```
library(skimr)

skim(mtcars)
```

Table: Data summary

```
Name          mtcars
Number of rows 32
Number of columns 11
-
Column type frequency:
numeric       11
-
Group variables  None
```

## Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
mpg	0	1	20.09	6.03	10.40	15.43	19.20	22.80	33.90	
cyl	0	1	6.19	1.79	4.00	4.00	6.00	8.00	8.00	
disp	0	1	230.72	123.94	71.10	120.83	196.30	326.00	472.00	
hp	0	1	146.69	68.56	52.00	96.50	123.00	180.00	335.00	
drat	0	1	3.60	0.53	2.76	3.08	3.70	3.92	4.93	
wt	0	1	3.22	0.98	1.51	2.58	3.33	3.61	5.42	
qsec	0	1	17.85	1.79	14.50	16.89	17.71	18.90	22.90	
vs	0	1	0.44	0.50	0.00	0.00	0.00	1.00	1.00	
am	0	1	0.41	0.50	0.00	0.00	0.00	1.00	1.00	
gear	0	1	3.69	0.74	3.00	3.00	4.00	4.00	5.00	
carb	0	1	2.81	1.62	1.00	2.00	2.00	4.00	8.00	

# A. Summarizing variables – skimr

Grouping also works with some of these summary functions

```
mtcars %>%  
  group_by(am) %>%  
  skim()
```

Table: Data summary

```
Name                Piped data  
Number of rows      32  
Number of columns   11  
-  
Column type frequency:  
numeric             10  
-  
Group variables     am
```

## Variable type: numeric

skim_variable	am	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
mpg	0	0	1	17.15	3.83	10.40	14.95	17.30	19.20	24.40	
mpg	1	0	1	24.39	6.17	15.00	21.00	22.80	30.40	33.90	
cyl	0	0	1	6.95	1.54	4.00	6.00	8.00	8.00	8.00	
cyl	1	0	1	5.08	1.55	4.00	4.00	4.00	6.00	8.00	
disp	0	0	1	290.38	110.17	120.10	196.30	275.80	360.00	472.00	
disp	1	0	1	143.53	87.20	71.10	79.00	120.30	160.00	351.00	
hp	0	0	1	160.26	53.91	62.00	116.50	175.00	192.50	245.00	
hp	1	0	1	126.85	84.06	52.00	66.00	109.00	113.00	335.00	
drat	0	0	1	3.29	0.39	2.76	3.07	3.15	3.70	3.92	
drat	1	0	1	4.05	0.36	3.54	3.85	4.08	4.22	4.93	
wt	0	0	1	3.77	0.78	2.46	3.44	3.52	3.84	5.42	

# A. Summarizing variables – psych

```
library(psych)
```

```
describe(mtcars)
```

```
##      vars  n   mean      sd median trimmed   mad   min   max   range  skew
## mpg      1 32  20.09    6.03  19.20   19.70    5.41 10.40  33.90  23.50  0.61
## cyl      2 32   6.19    1.79   6.00    6.23    2.97  4.00   8.00   4.00 -0.17
## disp     3 32 230.72 123.94 196.30  222.52  140.48 71.10 472.00 400.90  0.38
## hp       4 32 146.69  68.56 123.00  141.19   77.10 52.00 335.00 283.00  0.73
## drat     5 32   3.60    0.53   3.70    3.58    0.70  2.76   4.93   2.17  0.27
## wt       6 32   3.22    0.98   3.33    3.15    0.77  1.51   5.42   3.91  0.42
## qsec     7 32  17.85    1.79  17.71  17.83    1.42 14.50  22.90   8.40  0.37
## vs       8 32   0.44    0.50   0.00    0.42    0.00  0.00   1.00   1.00  0.24
## am       9 32   0.41    0.50   0.00    0.38    0.00  0.00   1.00   1.00  0.36
## gear    10 32   3.69    0.74   4.00    3.62    1.48  3.00   5.00   2.00  0.53
## carb    11 32   2.81    1.62   2.00    2.65    1.48  1.00   8.00   7.00  1.05
##      kurtosis      se
## mpg      -0.37  1.07
## cyl     -1.76  0.32
## disp    -1.21 21.91
## hp      -0.14 12.12
## drat    -0.71  0.09
## wt     -0.02  0.17
## qsec     0.34  0.32
## vs     -2.00  0.09
## am     -1.92  0.09
## gear    -1.07  0.13
## carb     1.26  0.29
```

# A. Summarizing variables

Can also use subsets of variables

```
mtcars %>% select(mpg, cyl, disp) %>%  
  describe()
```

```
##      vars  n   mean      sd median trimmed   mad  min   max range  skew  
## mpg      1 32  20.09   6.03   19.2   19.70   5.41 10.4  33.9  23.5  0.61  
## cyl      2 32   6.19   1.79    6.0    6.23   2.97  4.0   8.0   4.0 -0.17  
## disp     3 32 230.72 123.94  196.3  222.52 140.48 71.1 472.0 400.9  0.38  
##      kurtosis    se  
## mpg      -0.37  1.07  
## cyl      -1.76  0.32  
## disp     -1.21 21.91
```



# A. Summarizing variables

Or subsets of rows

```
mtcars %>% filter(str_detect(row.names(mtcars), "Merc")) %>%  
  describe()
```

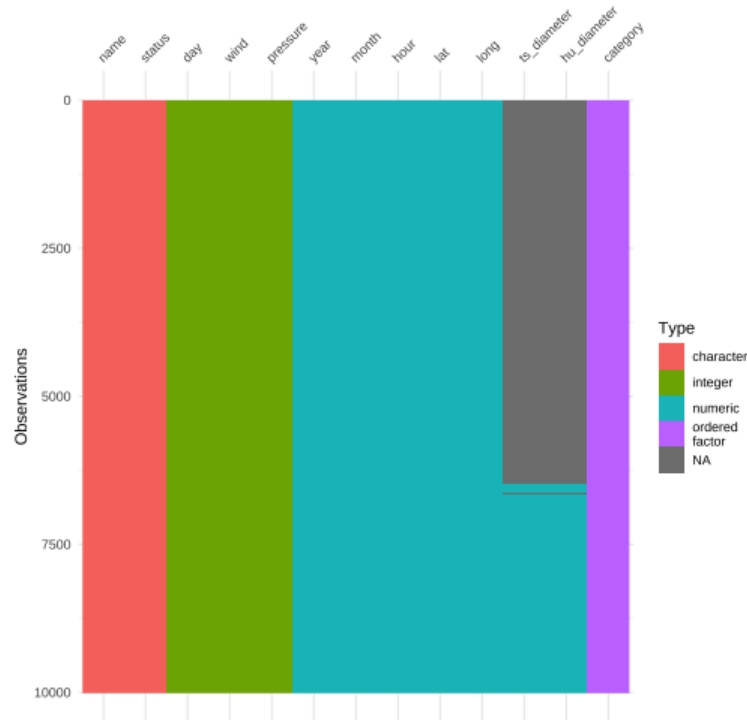
##	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew
##	mpg	1	7	19.01	3.40	17.80	19.01	2.08	15.20	24.40	9.20 0.48
##	cyl	2	7	6.29	1.80	6.00	6.29	2.97	4.00	8.00	4.00 -0.22
##	disp	3	7	207.16	64.97	167.60	207.16	39.73	140.80	275.80	135.00 0.17
##	hp	4	7	134.71	47.07	123.00	134.71	84.51	62.00	180.00	118.00 -0.24
##	drat	5	7	3.52	0.43	3.69	3.52	0.34	3.07	3.92	0.85 -0.15
##	wt	6	7	3.54	0.33	3.44	3.54	0.43	3.15	4.07	0.92 0.23
##	qsec	7	7	19.01	1.92	18.30	19.01	1.04	17.40	22.90	5.50 1.03
##	vs	8	7	0.57	0.53	1.00	0.57	0.00	0.00	1.00	1.00 -0.23
##	am	9	7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00 NaN
##	gear	10	7	3.57	0.53	4.00	3.57	0.00	3.00	4.00	1.00 -0.23
##	carb	11	7	3.00	0.82	3.00	3.00	1.48	2.00	4.00	2.00 0.00
##	kurtosis		se								
##	mpg	-1.59		1.28							
##	cyl	-1.90		0.68							
##	disp	-2.16		24.55							
##	hp	-1.75		17.79							
##	drat	-2.17		0.16							
##	wt	-1.61		0.13							
##	qsec	-0.49		0.73							
##	vs	-2.20		0.20							
##	am	NaN		0.00							
##	gear	-2.20		0.20							
##	carb	-1.71		0.31							

## B. Inspecting data frames and correlations

Another handy package is visdat: [https://cran.r-project.org/web/packages/visdat/vignettes/using\\_visdat.html](https://cran.r-project.org/web/packages/visdat/vignettes/using_visdat.html)

Lets you essentially visualize the sort of information you see in `glimpse()`

```
library(visdat)
vis_dat(storms)
```

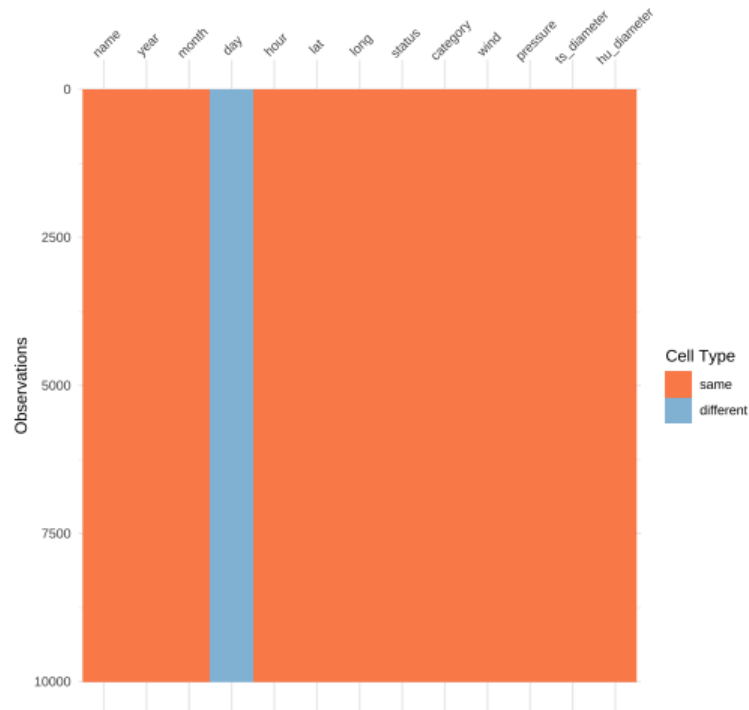


## B. Inspecting data frames and correlations

Another useful function from visdat is `vis_compare()`

This lets you compare two data sets of the same size to see what's different

```
storms_2 <- storms %>% mutate(day = as.numeric(day))  
vis_compare(storms, storms_2)
```



## B. Inspecting data frames and correlations

Correlations are something you'll probably want to look at.

Base R:

```
cor(mtcars)
```

```
##           mpg           cyl           disp           hp           drat           wt
## mpg      1.0000000 -0.8521620 -0.8475514 -0.7761684  0.68117191 -0.8676594
## cyl     -0.8521620  1.0000000  0.9020329  0.8324475 -0.69993811  0.7824958
## disp    -0.8475514  0.9020329  1.0000000  0.7909486 -0.71021393  0.8879799
## hp      -0.7761684  0.8324475  0.7909486  1.0000000 -0.44875912  0.6587479
## drat     0.6811719 -0.6999381 -0.7102139 -0.4487591  1.00000000 -0.7124406
## wt      -0.8676594  0.7824958  0.8879799  0.6587479 -0.71244065  1.0000000
## qsec     0.4186840 -0.5912421 -0.4336979 -0.7082234  0.09120476 -0.1747159
## vs       0.6640389 -0.8108118 -0.7104159 -0.7230967  0.44027846 -0.5549157
## am       0.5998324 -0.5226070 -0.5912270 -0.2432043  0.71271113 -0.6924953
## gear     0.4802848 -0.4926866 -0.5555692 -0.1257043  0.69961013 -0.5832870
## carb    -0.5509251  0.5269883  0.3949769  0.7498125 -0.09078980  0.4276059
##
##           qsec           vs           am           gear           carb
## mpg      0.41868403  0.6640389  0.59983243  0.4802848 -0.55092507
## cyl     -0.59124207 -0.8108118 -0.52260705 -0.4926866  0.52698829
## disp    -0.43369788 -0.7104159 -0.59122704 -0.5555692  0.39497686
## hp      -0.70822339 -0.7230967 -0.24320426 -0.1257043  0.74981247
## drat     0.09120476  0.4402785  0.71271113  0.6996101 -0.09078980
## wt      -0.17471588 -0.5549157 -0.69249526 -0.5832870  0.42760594
## qsec     1.00000000  0.7445354 -0.22986086 -0.2126822 -0.65624923
## vs       0.74453544  1.0000000  0.16834512  0.2060233 -0.56960714
## am      -0.22986086  0.1683451  1.00000000  0.7940588  0.05753435
## gear    -0.21268223  0.2060233  0.79405876  1.0000000  0.27407284
## carb    -0.65624923 -0.5696071  0.05753435  0.2740728  1.00000000
```

## B. Inspecting data frames and correlations

Base R:

Correlations are in a matrix object

```
storms %>%  
  select_if(is.numeric) %>%  
  cor()
```

```
##           year      month      day      hour      lat  
## year      1.000000000 -0.011488006  0.0183703369  0.0015741629 -0.121252667  
## month     -0.011488006  1.000000000 -0.1830702018 -0.0051201358 -0.065922836  
## day       0.018370337 -0.183070202  1.0000000000  0.0007164624 -0.050859874  
## hour      0.001574163 -0.005120136  0.0007164624  1.0000000000  0.002682367  
## lat      -0.121252667 -0.065922836 -0.0508598742  0.0026823666  1.000000000  
## long      0.060387523  0.048382680  0.0406477301 -0.0091876627 -0.104014683  
## wind      0.048966015  0.126682358 -0.0064971154  0.0018333102  0.076141764  
## pressure  -0.072615741 -0.134238300 -0.0010113895  0.0016030589 -0.103772744  
## ts_diameter      NA      NA      NA      NA      NA  
## hu_diameter      NA      NA      NA      NA      NA  
##           long      wind      pressure ts_diameter hu_diameter  
## year      0.060387523  0.048966015 -0.072615741      NA      NA  
## month     0.048382680  0.126682358 -0.134238300      NA      NA  
## day       0.040647730 -0.006497115 -0.001011389      NA      NA  
## hour      -0.009187663  0.001833310  0.001603059      NA      NA  
## lat      -0.104014683  0.076141764 -0.103772744      NA      NA  
## long      1.000000000  0.004737422  0.058467333      NA      NA  
## wind      0.004737422  1.000000000 -0.942249266      NA      NA  
## pressure  0.058467333 -0.942249266  1.000000000      NA      NA  
## ts_diameter      NA      NA      NA      1      NA  
## hu_diameter      NA      NA      NA      NA      1
```

## B. Inspecting data frames and correlations

Base R:

Correlations are in a matrix object

```
storms %>%  
  select_if(is.numeric) %>%  
  cor(use = "pairwise")
```

```
##           year      month      day      hour      lat  
## year      1.000000000 -0.011488006  0.0183703369  0.0015741629 -0.121252667  
## month     -0.011488006  1.000000000 -0.1830702018 -0.0051201358 -0.065922836  
## day       0.018370337 -0.183070202  1.0000000000  0.0007164624 -0.050859874  
## hour      0.001574163 -0.005120136  0.0007164624  1.0000000000  0.002682367  
## lat      -0.121252667 -0.065922836 -0.0508598742  0.0026823666  1.000000000  
## long      0.060387523  0.048382680  0.0406477301 -0.0091876627 -0.104014683  
## wind      0.048966015  0.126682358 -0.0064971154  0.0018333102  0.076141764  
## pressure  -0.072615741 -0.134238300 -0.0010113895  0.0016030589 -0.103772744  
## ts_diameter 0.021186700  0.139077211  0.0201075619  0.0085555295  0.300578521  
## hu_diameter -0.099658339  0.111651830  0.0338940169  0.0050504779  0.164416787  
##           long      wind      pressure ts_diameter hu_diameter  
## year      0.060387523  0.048966015 -0.072615741  0.021186700 -0.099658339  
## month     0.048382680  0.126682358 -0.134238300  0.139077211  0.111651830  
## day       0.040647730 -0.006497115 -0.001011389  0.020107562  0.033894017  
## hour      -0.009187663  0.001833310  0.001603059  0.008555529  0.005050478  
## lat      -0.104014683  0.076141764 -0.103772744  0.300578521  0.164416787  
## long      1.000000000  0.004737422  0.058467333 -0.014605508 -0.102351984  
## wind      0.004737422  1.000000000 -0.942249266  0.639640594  0.773608569  
## pressure  0.058467333 -0.942249266  1.000000000 -0.683340131 -0.842244047  
## ts_diameter -0.014605508  0.639640594 -0.683340131  1.000000000  0.683976179  
## hu_diameter -0.102351984  0.773608569 -0.842244047  0.683976179  1.000000000
```

## B. Inspecting data frames and correlations

corr package – `correlate()`

Introduces a new data frame type for correlations

```
library(corr)
```

```
correlate(mtcars)
```

```
## # A tibble: 11 x 12
##   term      mpg      cyl    disp      hp    drat      wt      qsec      vs      am
##   <chr>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1 mpg    NA      -0.852 -0.848 -0.776  0.681  -0.868  0.419  0.664  0.600
## 2 cyl   -0.852 NA       0.902  0.832 -0.700  0.782  -0.591 -0.811 -0.523
## 3 disp  -0.848 0.902 NA       0.791 -0.710  0.888  -0.434 -0.710 -0.591
## 4 hp    -0.776 0.832 0.791 NA      -0.449  0.659  -0.708 -0.723 -0.243
## 5 drat   0.681 -0.700 -0.710 -0.449 NA      -0.712  0.0912 0.440  0.713
## 6 wt    -0.868 0.782 0.888 0.659 -0.712 NA      -0.175 -0.555 -0.692
## 7 qsec   0.419 -0.591 -0.434 -0.708 0.0912 -0.175 NA      0.745 -0.230
## 8 vs     0.664 -0.811 -0.710 -0.723 0.440 -0.555 0.745 NA      0.168
## 9 am     0.600 -0.523 -0.591 -0.243 0.713 -0.692 -0.230 0.168 NA
## 10 gear  0.480 -0.493 -0.556 -0.126 0.700 -0.583 -0.213 0.206 0.794
## 11 carb -0.551 0.527 0.395 0.750 -0.0908 0.428 -0.656 -0.570 0.0575
## # ... with 2 more variables: gear <dbl>, carb <dbl>
```

## B. Inspecting data frames and correlations

corr package – Adds some useful features: variable subsets

```
mtcars %>% correlate() %>%  
  focus(mpg, cyl, disp)
```

```
## # A tibble: 8 x 4  
##   term      mpg    cyl  disp  
##   <chr>   <dbl> <dbl> <dbl>  
## 1 hp      -0.776  0.832  0.791  
## 2 drat     0.681 -0.700 -0.710  
## 3 wt       -0.868  0.782  0.888  
## 4 qsec     0.419 -0.591 -0.434  
## 5 vs        0.664 -0.811 -0.710  
## 6 am        0.600 -0.523 -0.591  
## 7 gear     0.480 -0.493 -0.556  
## 8 carb    -0.551  0.527  0.395
```



## B. Inspecting data frames and correlations

corr package – Adds some useful features: neater print output

```
mtcars %>% correlate() %>%  
  focus(mpg, cyl, disp) %>%  
  fashion()
```

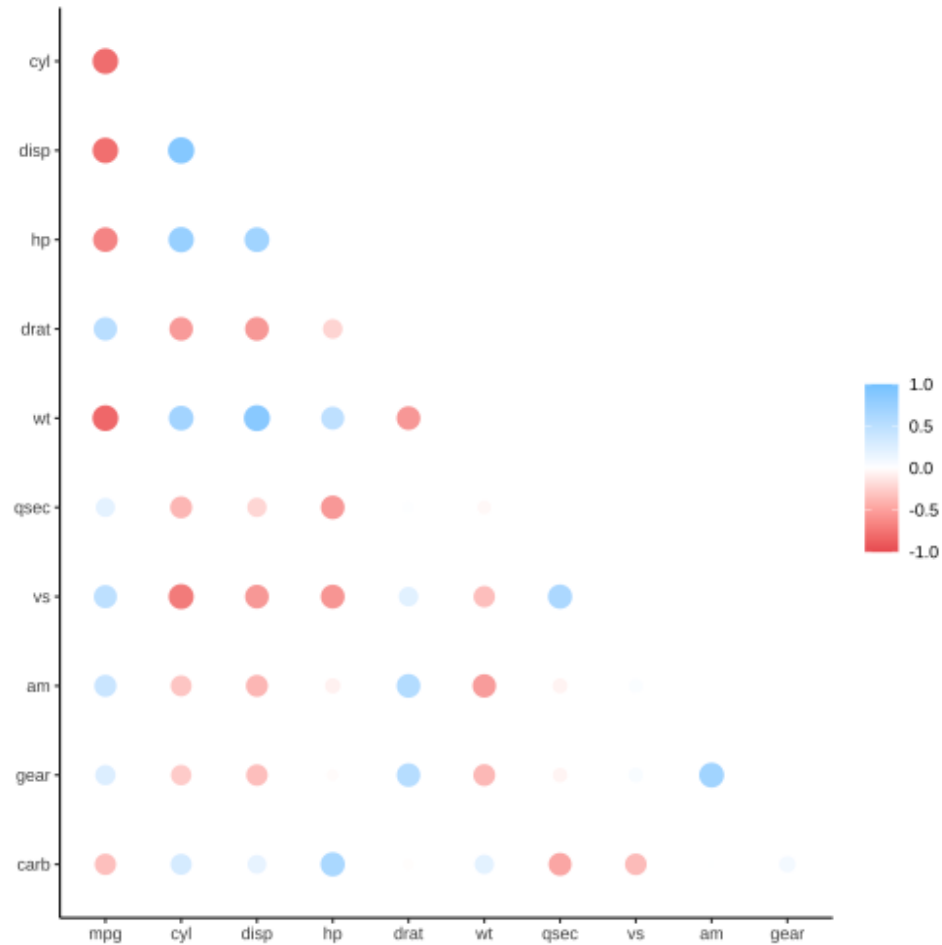
```
##   term  mpg  cyl disp  
## 1   hp  -.78  .83  .79  
## 2  drat  .68 -.70 -.71  
## 3   wt  -.87  .78  .89  
## 4  qsec  .42 -.59 -.43  
## 5   vs  .66 -.81 -.71  
## 6   am  .60 -.52 -.59  
## 7  gear  .48 -.49 -.56  
## 8  carb -.55  .53  .39
```

## B. Inspecting data frames and correlations

```
mtcars %>% correlate() %>%  
  rplot()
```

## B. Inspecting data frames and correlations

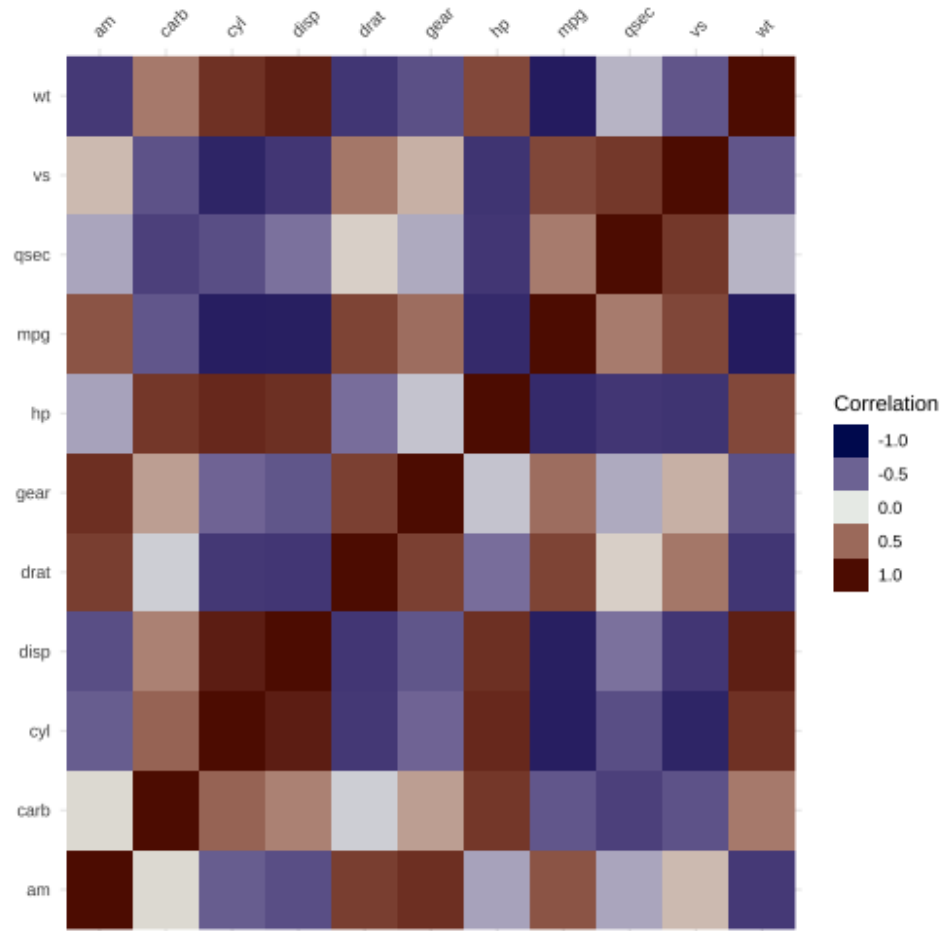
```
mtcars %>% correlate() %>%  
  shave() %>%  
  rplot()
```



## B. Inspecting data frames and correlations

visdat package does correlation heatmaps as well:

```
mtcars %>% vis_cor()
```



# C. Summary tables

We also can make some of this summary information neater for our Rmd documents

```
describe(mtcars)
```

```
##      vars  n   mean    sd median trimmed   mad   min    max   range  skew
## mpg      1 32  20.09   6.03  19.20   19.70   5.41 10.40  33.90  23.50  0.61
## cyl      2 32   6.19   1.79   6.00   6.23   2.97  4.00   8.00   4.00 -0.17
## disp     3 32 230.72 123.94 196.30  222.52 140.48 71.10 472.00 400.90  0.38
## hp       4 32 146.69  68.56 123.00  141.19  77.10 52.00 335.00 283.00  0.73
## drat     5 32   3.60   0.53   3.70   3.58   0.70  2.76   4.93   2.17  0.27
## wt       6 32   3.22   0.98   3.33   3.15   0.77  1.51   5.42   3.91  0.42
## qsec     7 32  17.85   1.79  17.71  17.83   1.42 14.50  22.90   8.40  0.37
## vs      8 32   0.44   0.50   0.00   0.42   0.00  0.00   1.00   1.00  0.24
## am      9 32   0.41   0.50   0.00   0.38   0.00  0.00   1.00   1.00  0.36
## gear    10 32   3.69   0.74   4.00   3.62   1.48  3.00   5.00   2.00  0.53
## carb    11 32   2.81   1.62   2.00   2.65   1.48  1.00   8.00   7.00  1.05
##      kurtosis    se
## mpg      -0.37  1.07
## cyl     -1.76  0.32
## disp     -1.21 21.91
## hp       -0.14 12.12
## drat     -0.71  0.09
## wt       -0.02  0.17
## qsec      0.34  0.32
## vs       -2.00  0.09
## am       -1.92  0.09
## gear     -1.07  0.13
## carb      1.26  0.29
```

## C. Summary tables

We also can make some of this summary information neater for our Rmd documents

```
library(knitr)
kable(describe(mtcars), digits = 2)
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
mpg	1	32	20.09	6.03	19.20	19.70	5.41	10.40	33.90	23.50	0.61	-0.37	1.07
cyl	2	32	6.19	1.79	6.00	6.23	2.97	4.00	8.00	4.00	-0.17	-1.76	0.32
disp	3	32	230.72	123.94	196.30	222.52	140.48	71.10	472.00	400.90	0.38	-1.21	21.91
hp	4	32	146.69	68.56	123.00	141.19	77.10	52.00	335.00	283.00	0.73	-0.14	12.12
drat	5	32	3.60	0.53	3.70	3.58	0.70	2.76	4.93	2.17	0.27	-0.71	0.09
wt	6	32	3.22	0.98	3.33	3.15	0.77	1.51	5.42	3.91	0.42	-0.02	0.17
qsec	7	32	17.85	1.79	17.71	17.83	1.42	14.50	22.90	8.40	0.37	0.34	0.32
vs	8	32	0.44	0.50	0.00	0.42	0.00	0.00	1.00	1.00	0.24	-2.00	0.09
am	9	32	0.41	0.50	0.00	0.38	0.00	0.00	1.00	1.00	0.36	-1.92	0.09
gear	10	32	3.69	0.74	4.00	3.62	1.48	3.00	5.00	2.00	0.53	-1.07	0.13
carb	11	32	2.81	1.62	2.00	2.65	1.48	1.00	8.00	7.00	1.05	1.26	0.29

## C. Summary tables

We also can make some of this summary information neater for our Rmd documents

```
cors <- mtcars %>% correlate() %>%  
  focus(mpg, cyl, disp) %>%  
  fashion()
```

```
cors
```

```
##   term mpg  cyl disp  
## 1  hp -.78  .83  .79  
## 2 drat .68 -.70 -.71  
## 3  wt -.87  .78  .89  
## 4 qsec .42 -.59 -.43  
## 5  vs .66 -.81 -.71  
## 6  am .60 -.52 -.59  
## 7 gear .48 -.49 -.56  
## 8 carb -.55  .53  .39
```

```
kable(cors)
```

```
term mpg cyl disp  
hp -.78 .83 .79  
drat .68 -.70 -.71  
wt -.87 .78 .89  
qsec .42 -.59 -.43  
vs .66 -.81 -.71  
am .60 -.52 -.59  
gear .48 -.49 -.56  
carb -.55 .53 .39
```

## C. Summary tables

We can also do some additional customizing with `kable()`

```
kable(describe(mtcars), digits = 2, caption = "This is my summary table")
```

Table: This is my summary table

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
mpg	1	32	20.09	6.03	19.20	19.70	5.41	10.40	33.90	23.50	0.61	-0.37	1.07
cyl	2	32	6.19	1.79	6.00	6.23	2.97	4.00	8.00	4.00	-0.17	-1.76	0.32
disp	3	32	230.72	123.94	196.30	222.52	140.48	71.10	472.00	400.90	0.38	-1.21	21.91
hp	4	32	146.69	68.56	123.00	141.19	77.10	52.00	335.00	283.00	0.73	-0.14	12.12
drat	5	32	3.60	0.53	3.70	3.58	0.70	2.76	4.93	2.17	0.27	-0.71	0.09
wt	6	32	3.22	0.98	3.33	3.15	0.77	1.51	5.42	3.91	0.42	-0.02	0.17
qsec	7	32	17.85	1.79	17.71	17.83	1.42	14.50	22.90	8.40	0.37	0.34	0.32
vs	8	32	0.44	0.50	0.00	0.42	0.00	0.00	1.00	1.00	0.24	-2.00	0.09
am	9	32	0.41	0.50	0.00	0.38	0.00	0.00	1.00	1.00	0.36	-1.92	0.09
gear	10	32	3.69	0.74	4.00	3.62	1.48	3.00	5.00	2.00	0.53	-1.07	0.13
carb	11	32	2.81	1.62	2.00	2.65	1.48	1.00	8.00	7.00	1.05	1.26	0.29



# C. Summary tables

The kableExtra packages adds more customization options to `kable()`

```
library(kableExtra)
table <- kable(describe(mtcars), digits = 2, caption = "This is my summary table")
row_spec(table, 5, bold = TRUE, background = "yellow") %>%
  footnote("only the highlighted row is important, but you can't have a table with just one row")
```

This is my summary table

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
mpg	1	32	20.09	6.03	19.20	19.70	5.41	10.40	33.90	23.50	0.61	-0.37	1.07
cyl	2	32	6.19	1.79	6.00	6.23	2.97	4.00	8.00	4.00	-0.17	-1.76	0.32
disp	3	32	230.72	123.94	196.30	222.52	140.48	71.10	472.00	400.90	0.38	-1.21	21.91
hp	4	32	146.69	68.56	123.00	141.19	77.10	52.00	335.00	283.00	0.73	-0.14	12.12
drat	5	32	3.60	0.53	3.70	3.58	0.70	2.76	4.93	2.17	0.27	-0.71	0.09
wt	6	32	3.22	0.98	3.33	3.15	0.77	1.51	5.42	3.91	0.42	-0.02	0.17
qsec	7	32	17.85	1.79	17.71	17.83	1.42	14.50	22.90	8.40	0.37	0.34	0.32
vs	8	32	0.44	0.50	0.00	0.42	0.00	0.00	1.00	1.00	0.24	-2.00	0.09
am	9	32	0.41	0.50	0.00	0.38	0.00	0.00	1.00	1.00	0.36	-1.92	0.09
gear	10	32	3.69	0.74	4.00	3.62	1.48	3.00	5.00	2.00	0.53	-1.07	0.13
carb	11	32	2.81	1.62	2.00	2.65	1.48	1.00	8.00	7.00	1.05	1.26	0.29

Note:

only the highlighted row is important, but you can't have a table with just one row

## C. Summary tables

There are other packages that can customize tables for specific formats

the sjPlot package is particularly useful for producing orderly and nice looking output tables for different kinds of models.

Example: `lm()` summary output

```
library(sjPlot)
model <- lm(mpg ~ disp + cyl, data = mtcars)
```

## C. Summary tables

Example: `lm()` print output

```
model <- lm(mpg ~ disp + cyl, data = mtcars)
model
```

```
##
## Call:
## lm(formula = mpg ~ disp + cyl, data = mtcars)
##
## Coefficients:
## (Intercept)      disp      cyl
##   34.66099   -0.02058   -1.58728
```

# C. Summary tables

Example: `lm()` summary output

```
model <- lm(mpg ~ disp + cyl, data = mtcars)
summary(model)
```

```
##
## Call:
## lm(formula = mpg ~ disp + cyl, data = mtcars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.4213 -2.1722 -0.6362  1.1899  7.0516
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  34.66099    2.54700   13.609 4.02e-14 ***
## disp        -0.02058    0.01026   -2.007  0.0542  .
## cyl         -1.58728    0.71184   -2.230  0.0337  *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.055 on 29 degrees of freedom
## Multiple R-squared:  0.7596,    Adjusted R-squared:  0.743
## F-statistic: 45.81 on 2 and 29 DF,  p-value: 1.058e-09
```

# C. Summary tables

Example: `lm()` sjPlots table output

```
model <- lm(mpg ~ disp + cyl, data = mtcars)
tab_model(model)
```

<i>Predictors</i>	<i>mpg</i>		
	<i>Estimates</i>	<i>CI</i>	<i>p</i>
(Intercept)	34.66	29.45 – 39.87	<b>&lt;0.001</b>
disp	–0.02	–0.04 – 0.00	0.054
cyl	–1.59	–3.04 – –0.13	<b>0.034</b>
Observations	32		
R <sup>2</sup> / R <sup>2</sup> adjusted	0.760 / 0.743		

## C. Summary tables

```
library(apaTables)  
apa.cor.table(attitude)  
apa.cor.table(attitude, filename="ex-CorTable1.doc")
```

# Live coding

Let's head over to the following file for a demonstration: Week 8 demo

# Feedback

---

*We're at the middle of the semester*

1. What should we keep doing?
2. What should we do more?
3. What should we do less?

Please consider responding this short mid-semester survey/feedback form (completely anonymous): <https://forms.gle/wuHkb7TgeZs7HoXd8>



# Logistics

---

## This week

- Homework 8: Available tomorrow by noon; **Due by Thursday, 3/18**
- Readings <https://r4ds.had.co.nz/exploratory-data-analysis.html>
- Schedule updated: <https://making-data-science-count.github.io/s21-intro-to-data-sci-methods-in-ed/schedule.html>

# Final Project

---

- Final project

# Wrapping up

In your base group's Slack channel:

- What is one thing you learned today?
- What is something you want to learn more about?
- Share your feelings in GIF form!

We really appreciate being able to see these reactions and get this feedback!