

Get links

```
bow(url = "https://www.imdb.com/search/title/?groups=b
scrape() %>%
html_elements(css = ".lister-item-header a") %>%
html_attr(name = "href") %>%
url_absolute(base = "https://www.imdb.com")
```

```
[1] "https://www.imdb.com/title/tt10366460/?ref=adv_
[2] "https://www.imdb.com/title/tt9770150/?ref=adv_l
[3] "https://www.imdb.com/title/tt6751668/?ref=adv_l
[4] "https://www.imdb.com/title/tt6966692/?ref=adv_l
[5] "https://www.imdb.com/title/tt5580390/?ref=adv_l
[6] "https://www.imdb.com/title/tt4975722/?ref=adv_l
[7] "https://www.imdb.com/title/tt1895587/?ref=adv_l
[8] "https://www.imdb.com/title/tt2562232/?ref=adv_l
[9] "https://www.imdb.com/title/tt2024544/?ref=adv_l
[10] "https://www.imdb.com/title/tt1024648/?ref=adv_l
[11] "https://www.imdb.com/title/tt1655442/?ref=adv_l
[12] "https://www.imdb.com/title/tt1504320/?ref=adv_l
[13] "https://www.imdb.com/title/tt1010048/?ref=adv_l
[14] "https://www.imdb.com/title/tt0887912/?ref=adv_l
[15] "https://www.imdb.com/title/tt0477348/?ref=adv_l
[16] "https://www.imdb.com/title/tt0407887/?ref=adv_l
[17] "https://www.imdb.com/title/tt0405159/?ref=adv_l
[18] "https://www.imdb.com/title/tt0375679/?ref=adv_l
[19] "https://www.imdb.com/title/tt0167260/?ref=adv_l
[20] "https://www.imdb.com/title/tt0299658/?ref=adv_l
[21] "https://www.imdb.com/title/tt0268978/?ref=adv_l
[22] "https://www.imdb.com/title/tt0172495/?ref=adv_l
[23] "https://www.imdb.com/title/tt0169547/?ref=adv_l
[24] "https://www.imdb.com/title/tt0138097/?ref=adv_l
[25] "https://www.imdb.com/title/tt0120338/?ref=adv_l
[26] "https://www.imdb.com/title/tt0116209/?ref=adv_l
[27] "https://www.imdb.com/title/tt0112573/?ref=adv_l
[28] "https://www.imdb.com/title/tt0109830/?ref=adv_l
[29] "https://www.imdb.com/title/tt0108052/?ref=adv_l
[30] "https://www.imdb.com/title/tt0105695/?ref=adv_l
```

Scrape table

```
table_usafacts <- bow(url = "https://usafacts.org/visualizations/covid-vaccine-tracker-state
  scrape() %>% html_elements(css = "table") %>% html_table() %>% pluck(1)
knitr::kable(table_usafacts, format = "html")
```

State	% of population with at least one dose	% fully vaccinated	% with booster or additional dose
Alabama	63.9%	52.2%	19.7%
Alaska	71.5%	63.4%	29.9%
Arizona	75.5%	63.1%	28.5%
Arkansas	68.2%	55.6%	23.4%
California	84.1%	73.5%	41%
Colorado	81.4%	71.6%	39.6%
Connecticut	>95%*	80.8%	43.3%
Delaware	85.3%	70.9%	34.4%
District of Columbia	>95%*	78.8%	34.8%
Florida	80.9%	68.2%	28.9%

Click [here](#) to take a look at the webpage

Scraping multiple tables

```
all_url <- "https://finance.yahoo.com/screener/predefined/day_gainers?count=25&offset="
idx <- seq(0, 1050, by = 25)

table_new <- data.frame()
df <- data.frame()

for (i in seq_along(idx)) {
  new_webpage <- read_html(str_glue(all_url, {idx[i]})) # same as bow(url) %>% scrape()
  table_new <- html_table(new_webpage)[[1]] %>% # first element of the list
    as_tibble(.name_repair = "unique") # repairs same column names
  df <- rbind(df, table_new)
}
```

Multiple tables combined

Show 5 ▾ entries		Search: <input type="text"/>							
	Symbol ▾	Name ▾	Price (Intraday) ▾	Change ▾	% Change ▾	Volume ▾	Avg Vol (3 month) ▾	Market Cap ▾	PE Ratio (TTM) ▾
1	LTMAQ	LATAM Airlines Group S.A.	0.364	0.0388	+11.93%	128,249	1.102M	3.957B	N/A
2	BANF	BancFirst Corporation	100.96	9.74	+10.67%	31,513	146,421	3.419B	21.34
3	SAM	The Boston Beer Company, Inc.	370.4	34.24	+10.19%	154,134	146,862	4.542B	N/A
4	SMPL	The Simply Good Foods Company	35.26	2.88	+8.89%	334,694	687,798	3.589B	37.51
5	HBAN	Huntington Bancshares Incorporated	14.31	1.11	+8.45%	10.378M	15.99M	21.302B	12.56
Showing 1 to 5 of 107 entries			Previous 1 2 3 4 5 ... 22 Next						

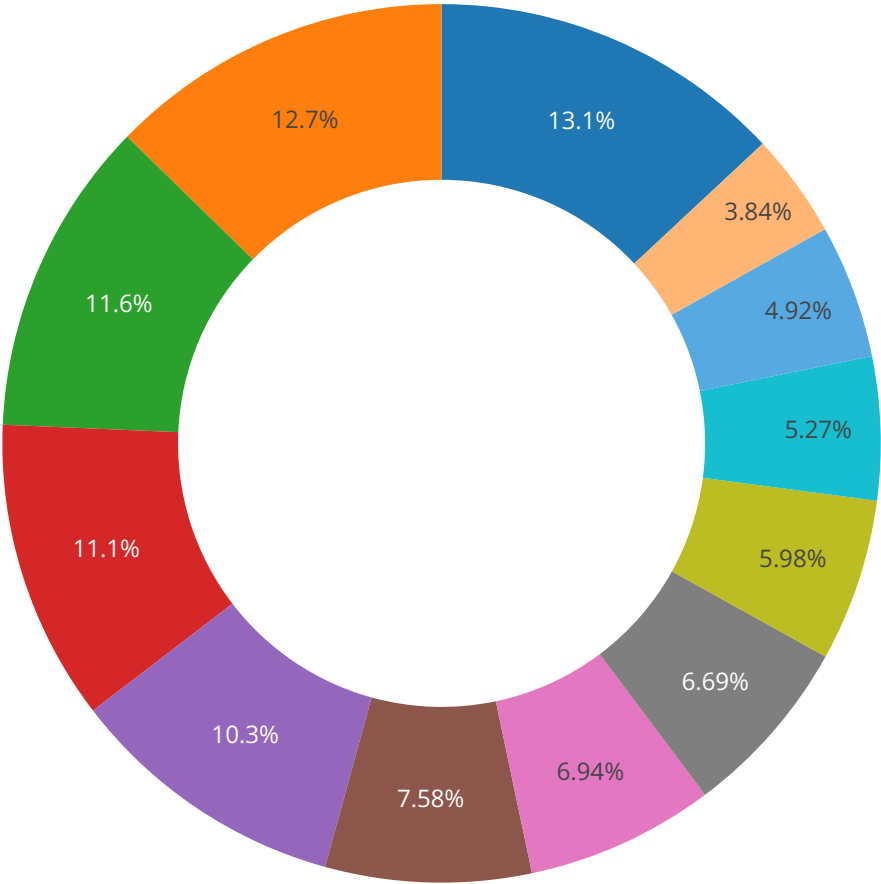
Tidy further

```
df_movies %>%  
  rename(ID = `...1`) %>%  
  mutate(ProductionBudget = parse_number(ProductionBudget),  
         DomesticGross = parse_number(DomesticGross),  
         WorldwideGross = parse_number(WorldwideGross),  
         ReleaseDate = mdy(ReleaseDate),  
         ReleaseDate = replace_na(ReleaseDate, make_date()),  
         MonthOfRelease = month(ReleaseDate, label = TRUE),  
         YearOfRelease = year(ReleaseDate)) %>%  
  select(MonthOfRelease, DomesticGross) %>%  
  group_by(MonthOfRelease) %>%  
  summarize(AverageByMonth = mean(DomesticGross))
```

```
# A tibble: 12 × 2  
  MonthOfRelease AverageByMonth  
  <ord>          <dbl>  
1 Jan           19168511.  
2 Feb           34627213.  
3 Mar           37815136.  
4 Apr           33408498.  
5 May           63397468.  
6 Jun           65140937.  
7 Jul           55330277.  
8 Aug           29849508.  
9 Sep           24579997.  
10 Oct          26305125.  
11 Nov          51487619.  
12 Dec          58006543.
```

Interactive Donut Plot

Average Domestic Gross by Month



Interactive visualizations using Plotly

```
midwest %>% as_tibble()
```

```
# A tibble: 437 × 28
```

	PID	county	state	area	poptotal	popden... ¹	popwh... ²	popbl... ³	popam... ⁴	popas... ⁵
	<int>	<chr>	<chr>	<dbl>	<int>	<dbl>	<int>	<int>	<int>	<int>
1	561	ADAMS	IL	0.052	66090	1271.	63917	1702	98	249
2	562	ALEXANDER	IL	0.014	10626	759	7054	3496	19	48
3	563	BOND	IL	0.022	14991	681.	14477	429	35	16
4	564	BOONE	IL	0.017	30806	1812.	29344	127	46	150
5	565	BROWN	IL	0.018	5836	324.	5264	547	14	5
6	566	BUREAU	IL	0.05	35688	714.	35157	50	65	195
7	567	CALHOUN	IL	0.017	5322	313.	5298	1	8	15
8	568	CARROLL	IL	0.027	16805	622.	16519	111	30	61
9	569	CASS	IL	0.024	13437	560.	13384	16	8	23
10	570	CHAMPAIGN	IL	0.058	172025	2082	146506	16550	221	8022

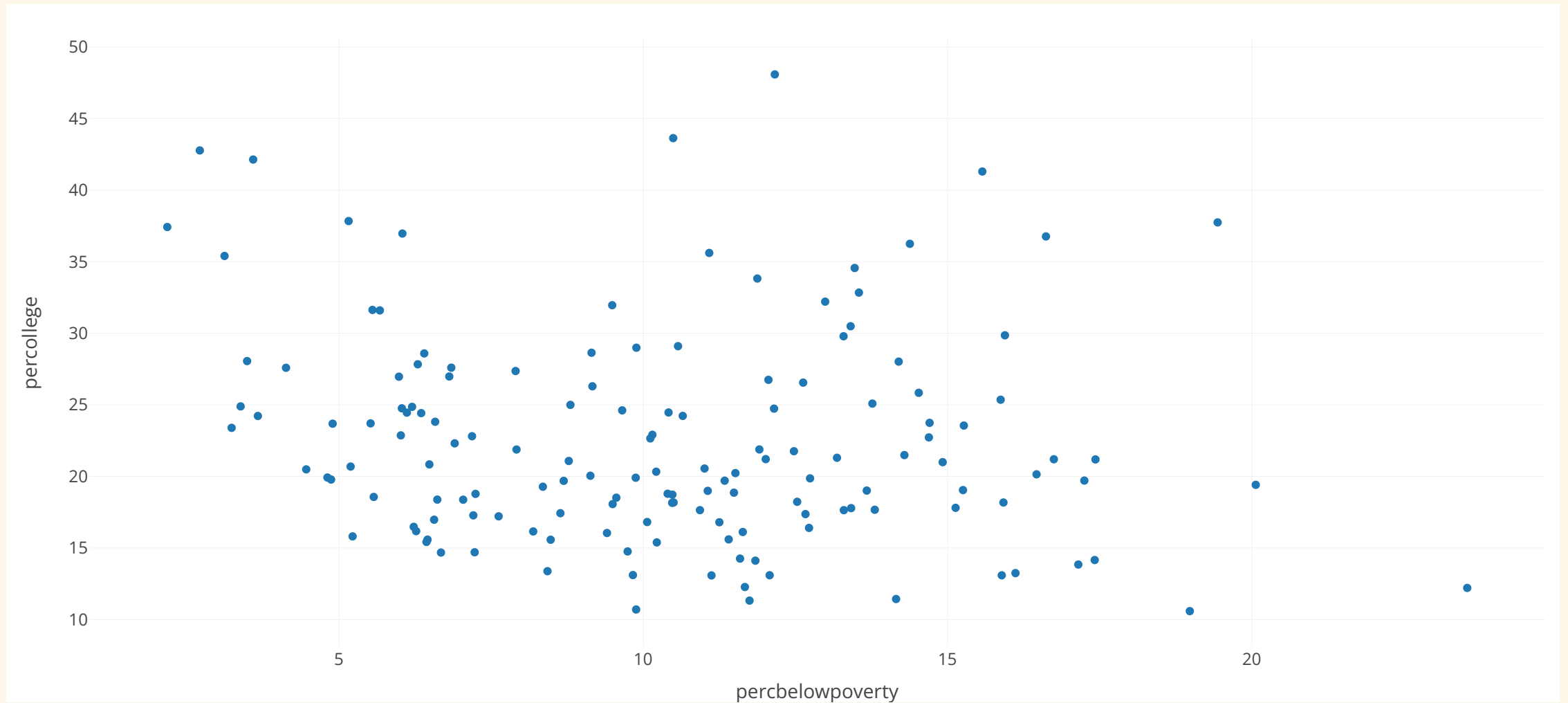
```
midwest %>%
```

```
  filter(inmetro == T) %>%
```

```
  plot_ly(x = ~ percbelowpoverty, y = ~ percollege) %>%
```

```
  add_markers()
```

Interactive visualizations using Plotly



Interactive visualizations using ggplotly

```
mtcars %>% as_tibble() %>% head()
```

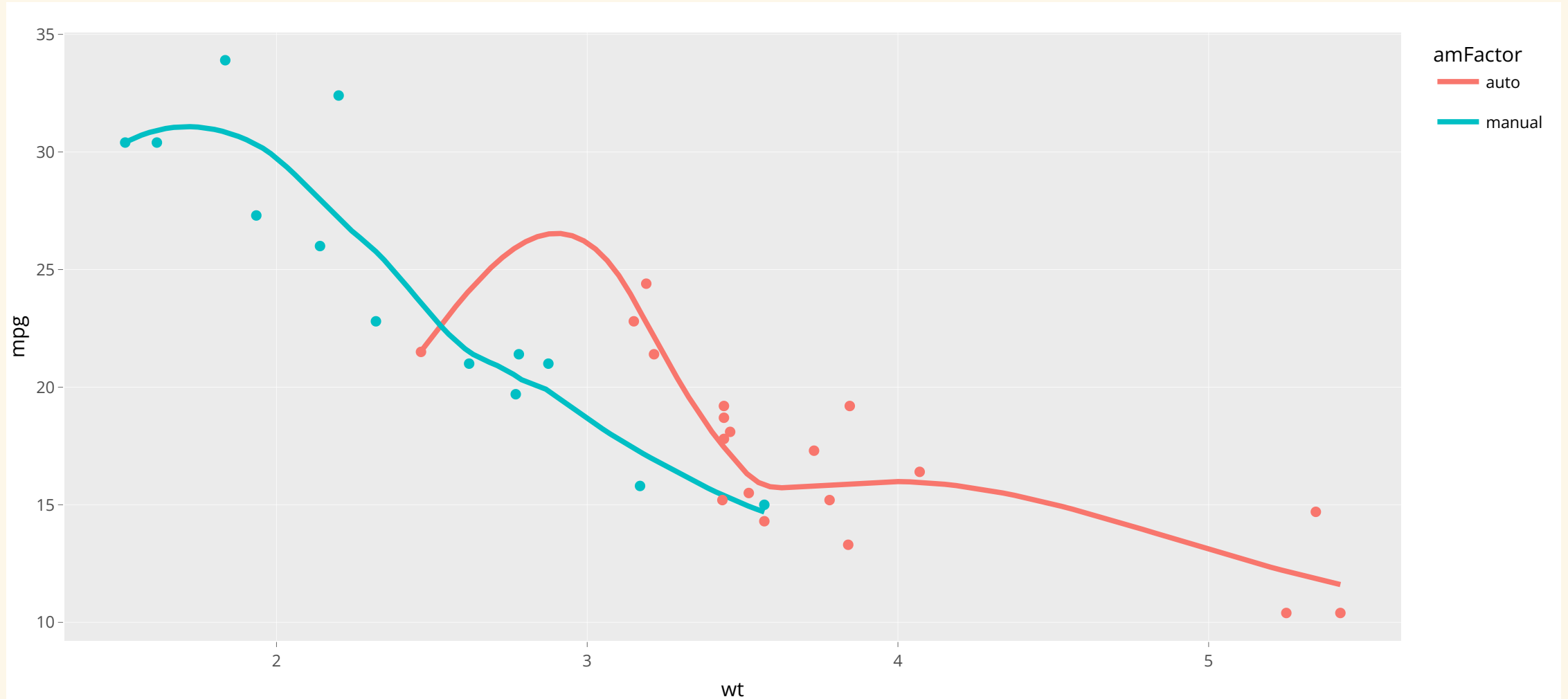
```
# A tibble: 6 × 11
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	21	6	160	110	3.9	2.62	16.5	0	1	4	4
2	21	6	160	110	3.9	2.88	17.0	0	1	4	4
3	22.8	4	108	93	3.85	2.32	18.6	1	1	4	1
4	21.4	6	258	110	3.08	3.22	19.4	1	0	3	1
5	18.7	8	360	175	3.15	3.44	17.0	0	0	3	2
6	18.1	6	225	105	2.76	3.46	20.2	1	0	3	1

```
gp = mtcars %>%  
  mutate(amFactor = factor(am, labels = c('auto', 'manual')),  
         hovertext = paste(wt, mpg, amFactor)) %>%  
  arrange(wt) %>%  
  ggplot(aes(x = wt, y = mpg, color = amFactor)) +  
  geom_smooth(se = F) +  
  geom_point(aes(color = amFactor)) + theme_economist_white()
```

```
ggplotly()
```

Interactive visualizations using `ggplotly`



DT: Interactive Data Tables

```
library(ggplot2movies)
movies %>%
  select(1:6) %>%
  filter(rating > 8, !is.na(budget), votes > 1000) %>%
  datatable(fillContainer = FALSE, options = list(pageLength = 6))
```

Show 6 entries

Search:

	title	year	length	budget	rating	votes
1	12 Angry Men	1957	96	340000	8.7	29278
2	2001: A Space Odyssey	1968	156	10500000	8.3	64982
3	Adventures of Robin Hood, The	1938	102	1900000	8.2	7359
4	Alien	1979	116	11000000	8.3	63400
5	Aliens	1986	154	18500000	8.3	63961
6	All Quiet on the Western Front	1930	147	1200000	8.2	6835

Showing 1 to 6 of 149 entries

Previous

1

2

3

4

5

...

25

Next

Group Activity 2

10:00



- Work on activity 2
- Ask me questions