

# AMOUNTS

MPA 635: Data Visualization

September 25, 2018

# PLAN FOR TODAY

---

More on truth

Amounts

Verbs

Live example

# MORE ON TRUTH

# DATA AND WHITE LIES

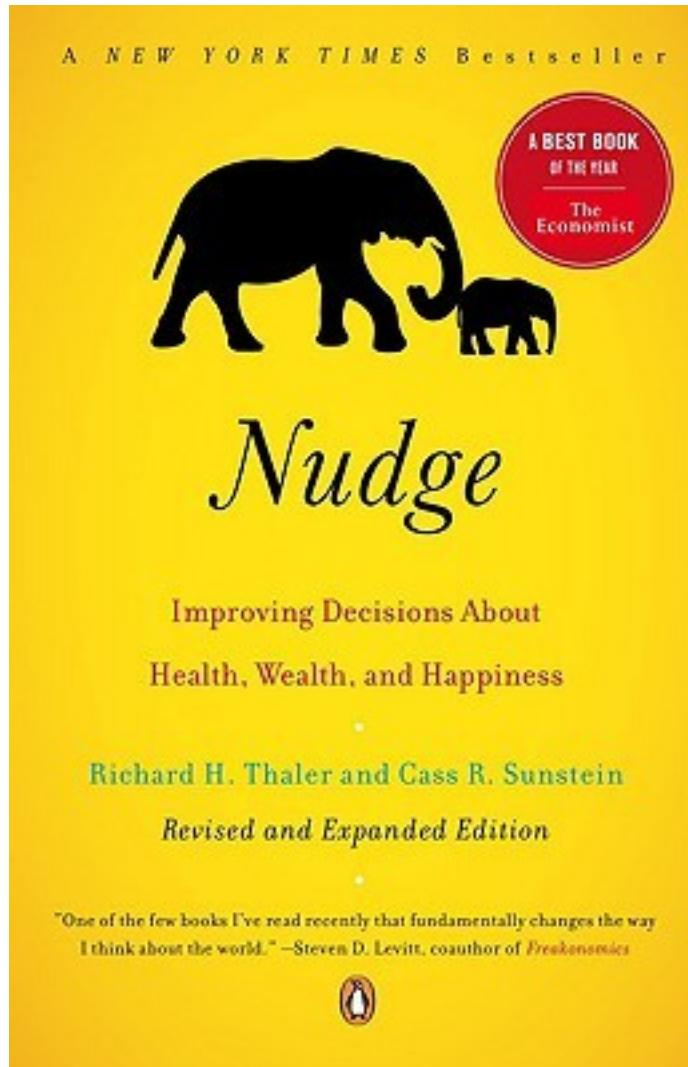
---

“I secretly wonder if I'm a righteous dude, is it OK for me to sort of maybe possibly mislead people so they pursue a more righteous policy?”

Anonymous MPA 635 student

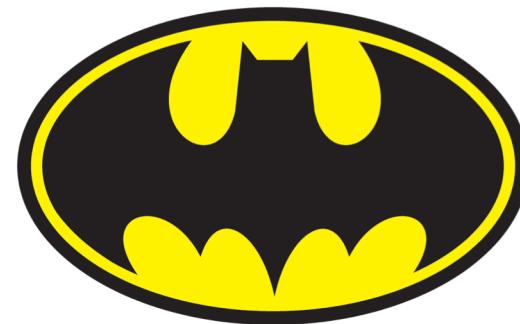
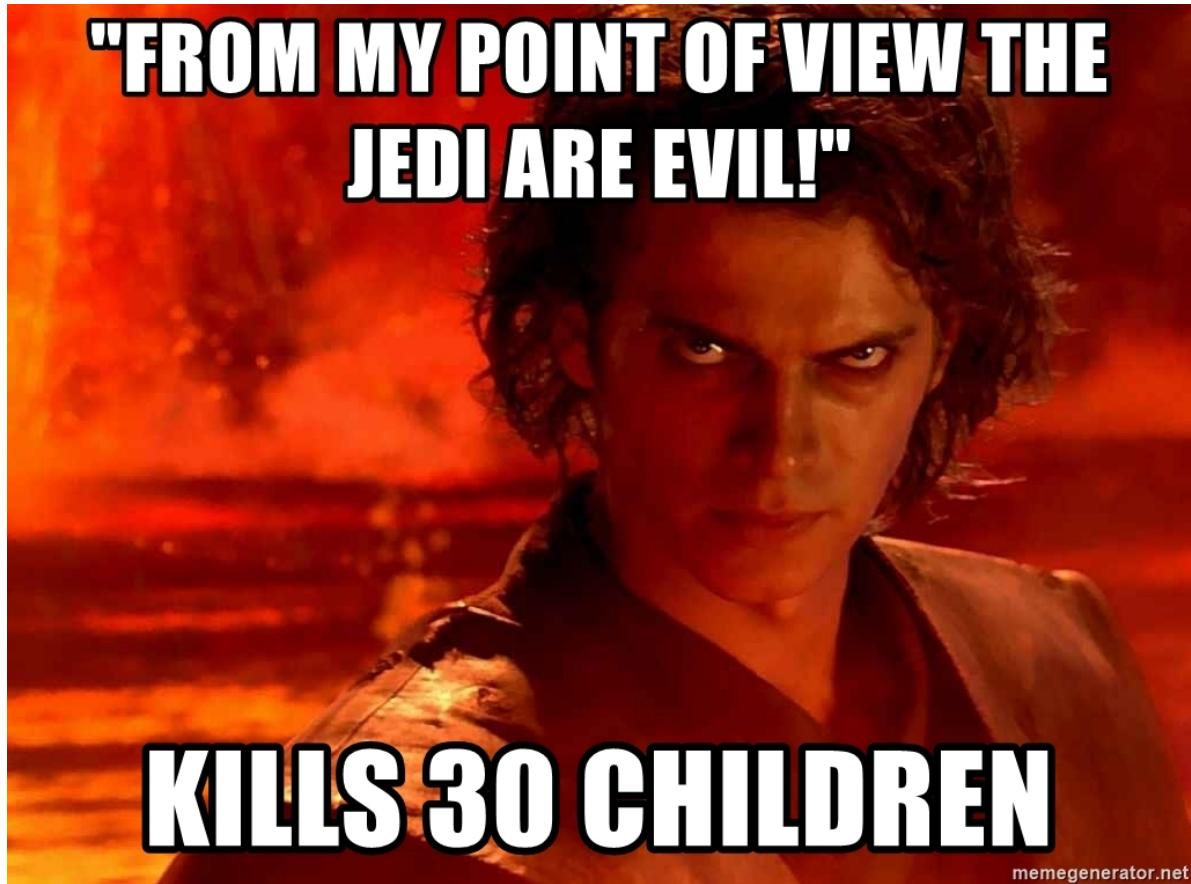
# IS NUDGING OKAY?

---



# WHO DEFINES "GOOD"?

---



# DON'T MESS WITH DATA

You can push people towards policy outcomes, but don't distort data to do it.

“Lies, damned lies, and statistics”

↑ Don't perpetuate this ↑

# AMOUNTS

# PROBLEMS WITH BAR PLOTS

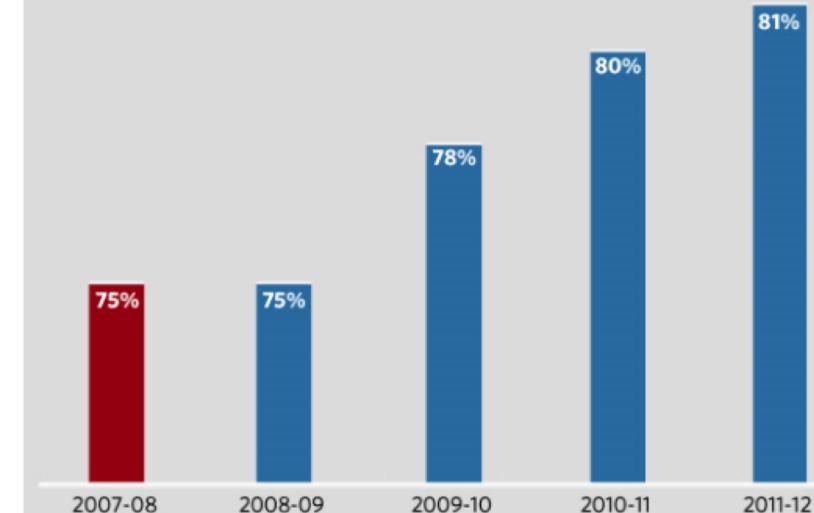
---



## CHART OF THE WEEK

Under President Obama,  
**OUR PUBLIC HIGH SCHOOL GRADUATION RATE IS THE HIGHEST IT'S EVER BEEN**

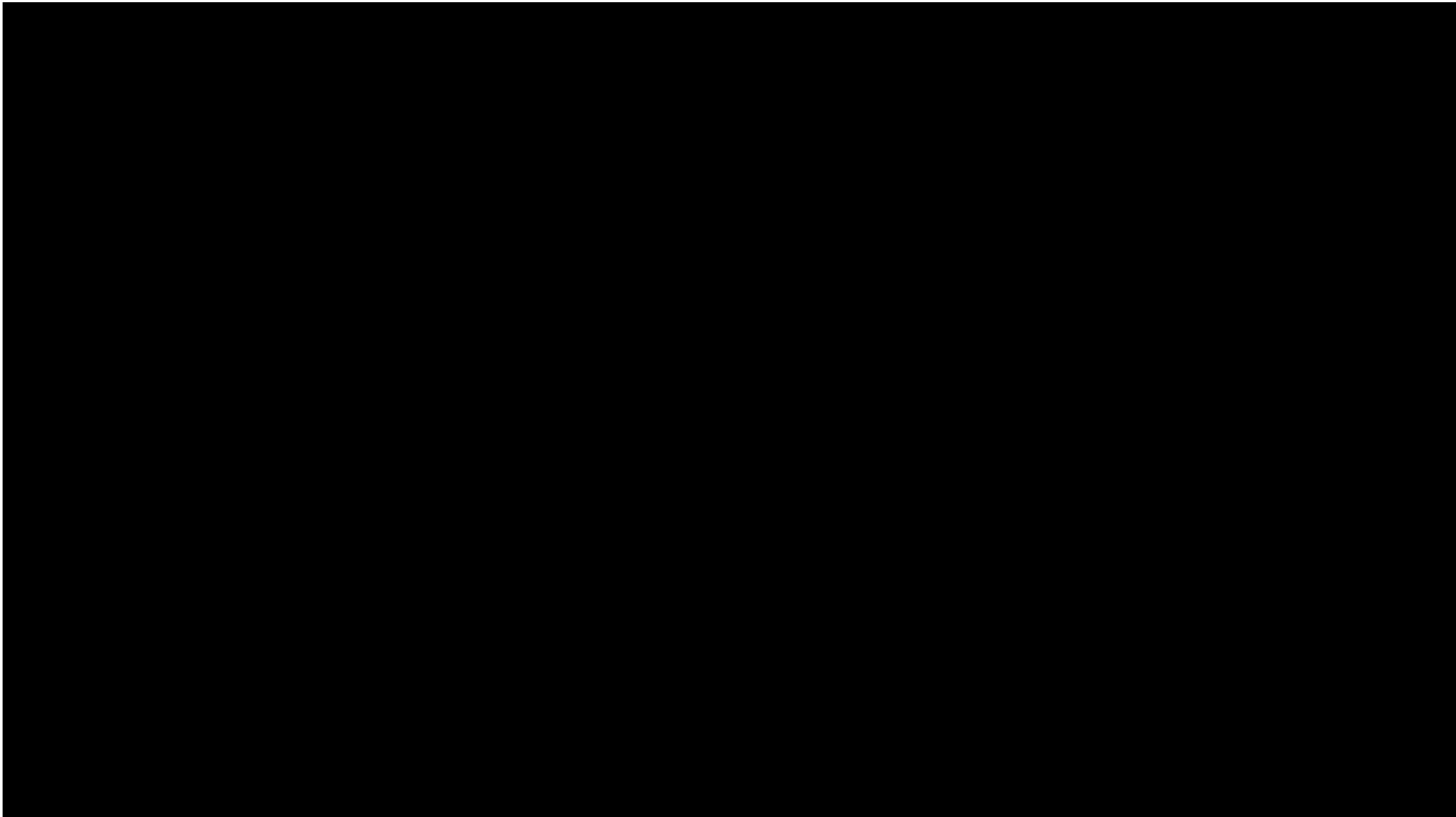
AVERAGED FRESHMEN GRADUATION RATE



Source: U.S. Department of Education, National Center for Education Statistics

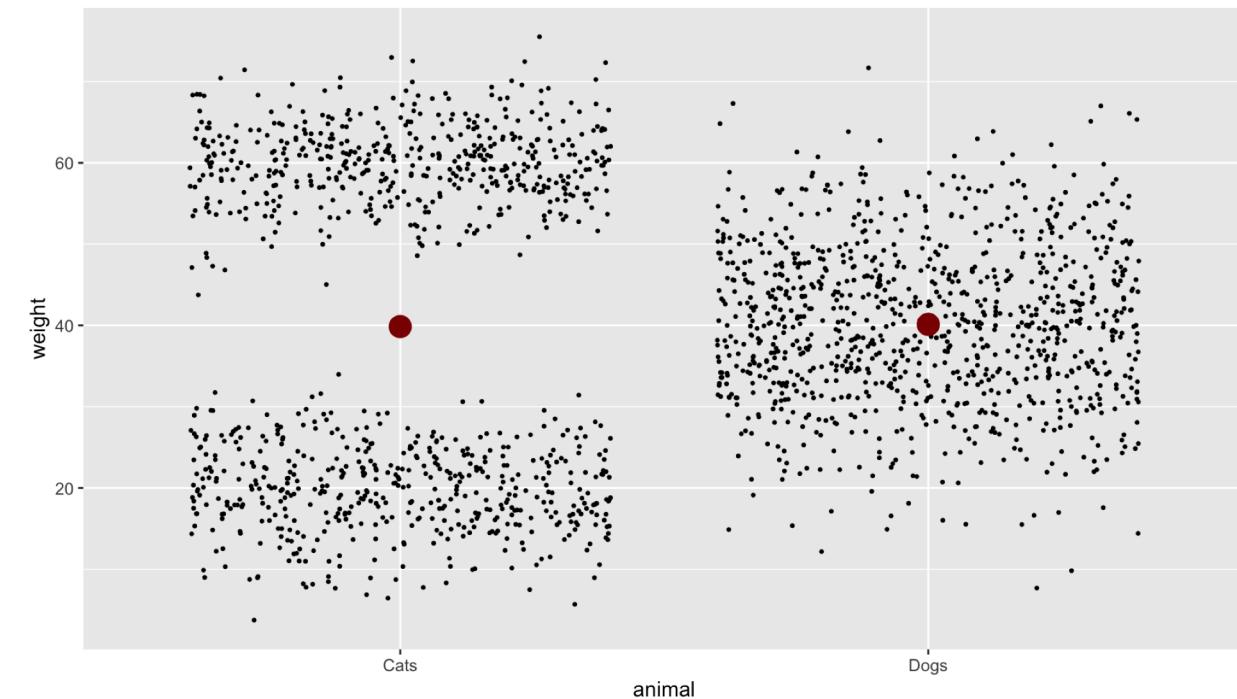
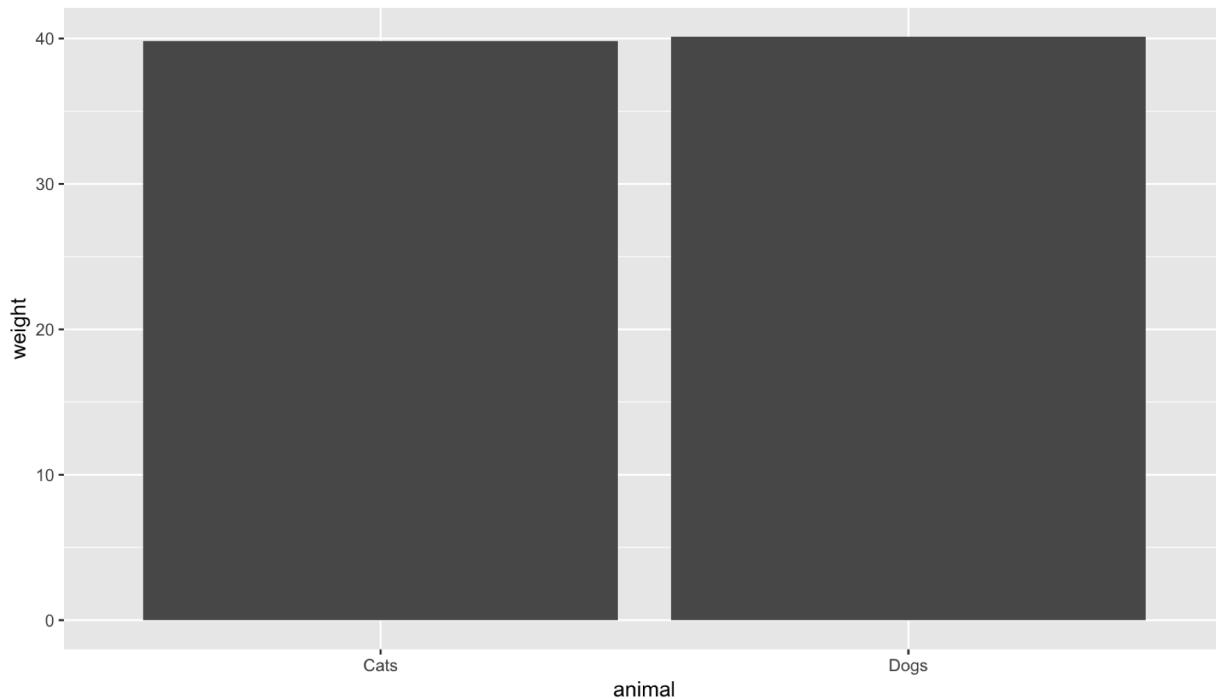
# # bar plots

---



# BAR PLOTS AND SUMMARY STATS

---



# GENERAL RULES

---

**More data = better**

**Don't use bars for  
summary stats**

**The end of the bar is  
often all that matters**

Show actual points

Counts okay, but there  
are better solutions

Lollipops, points,  
heatmaps

**Always start at zero!**

# VERBS

# MOST COMMON VERBS

---

`filter()`

Choose rows based on conditions

`select()`

Choose (and rename) columns

`mutate()`

Add column (or change existing column)

`group_by()`

Make subgroups based on a column

`summarize()`

Calculate summary statistics for groups

# FILTER

---

```
gapminder %>%  
  filter(year == 1967)
```

country <fctr>	continent <fctr>	year <int>	lifeExp <dbl>	pop <int>	gdpPerCap <dbl>
Afghanistan	Asia	1967	34.02000	11537966	836.1971
Albania	Europe	1967	66.22000	1984060	2760.1969
Algeria	Africa	1967	51.40700	12760499	3246.9918
Angola	Africa	1967	35.98500	5247469	5522.7764
Argentina	Americas	1967	65.63400	22934225	8052.9530
Australia	Oceania	1967	71.10000	11872264	14526.1246
Austria	Europe	1967	70.14000	7376998	12834.6024
Bahrain	Asia	1967	59.92300	202182	14804.6727
Bangladesh	Asia	1967	43.45300	62821884	721.1861
Belgium	Europe	1967	70.94000	9556500	13149.0412

1-10 of 142 rows

Previous 1 2 3 4 5 6 ... 15 Next

# FILTER

---

```
gapminder %>%  
  filter(lifeExp < 40)
```

country <fctr>	continent <fctr>	year <int>	lifeExp <dbl>	pop <int>	gdpPerCap <dbl>
Afghanistan	Asia	1952	28.801	8425333	779.4453
Afghanistan	Asia	1957	30.332	9240934	820.8530
Afghanistan	Asia	1962	31.997	10267083	853.1007
Afghanistan	Asia	1967	34.020	11537966	836.1971
Afghanistan	Asia	1972	36.088	13079460	739.9811
Afghanistan	Asia	1977	38.438	14880372	786.1134
Afghanistan	Asia	1982	39.854	12881816	978.0114
Angola	Africa	1952	30.015	4232095	3520.6103
Angola	Africa	1957	31.999	4561361	3827.9405
Angola	Africa	1962	34.000	4826015	4269.2767

1-10 of 124 rows

Previous 1 2 3 4 5 6 ... 13 Next

# FILTER

---

```
gapminder %>%  
  filter(continent == "Asia", lifeExp < 40)
```

country <fctr>	continent <fctr>	year <int>	lifeExp <dbl>	pop <int>	gdpPerCap <dbl>
Afghanistan	Asia	1952	28.801	8425333	779.4453
Afghanistan	Asia	1957	30.332	9240934	820.8530
Afghanistan	Asia	1962	31.997	10267083	853.1007
Afghanistan	Asia	1967	34.020	11537966	836.1971
Afghanistan	Asia	1972	36.088	13079460	739.9811
Afghanistan	Asia	1977	38.438	14880372	786.1134
Afghanistan	Asia	1982	39.854	12881816	978.0114
Bangladesh	Asia	1952	37.484	46886859	684.2442
Bangladesh	Asia	1957	39.348	51365468	661.6375
Cambodia	Asia	1952	39.417	4693836	368.4693

1-10 of 25 rows

Previous 1 2 3 Next

# SELECT

---

```
gapminder %>%  
  select(country, year, pop)
```

country <fctr>	year <int>	pop <int>
Afghanistan	1952	8425333
Afghanistan	1957	9240934
Afghanistan	1962	10267083
Afghanistan	1967	11537966
Afghanistan	1972	13079460
Afghanistan	1977	14880372
Afghanistan	1982	12881816
Afghanistan	1987	13867957
Afghanistan	1992	16317921
Afghanistan	1997	22227415

1-10 of 1,704 rows

Previous 1 2 3 4 5 6 ... 100 Next

# MUTATE

---

```
gapminder %>%  
  mutate(something_new = 5)
```

country <fctr>	continent <fctr>	year <int>	lifeExp <dbl>	pop <int>	gdpPercap <dbl>	something_new <dbl>
Afghanistan	Asia	1952	28.80100	8425333	779.4453	5
Afghanistan	Asia	1957	30.33200	9240934	820.8530	5
Afghanistan	Asia	1962	31.99700	10267083	853.1007	5
Afghanistan	Asia	1967	34.02000	11537966	836.1971	5
Afghanistan	Asia	1972	36.08800	13079460	739.9811	5
Afghanistan	Asia	1977	38.43800	14880372	786.1134	5
Afghanistan	Asia	1982	39.85400	12881816	978.0114	5
Afghanistan	Asia	1987	40.82200	13867957	852.3959	5
Afghanistan	Asia	1992	41.67400	16317921	649.3414	5
Afghanistan	Asia	1997	41.76300	22227415	635.3414	5

1-10 of 1,704 rows

Previous 1 2 3 4 5 6 ... 100 Next

# MUTATE

---

```
gapminder %>%  
  mutate(pop_million = pop / 1000000)
```

country <fctr>	continent <fctr>	year <int>	lifeExp <dbl>	pop <int>	gdpPercap <dbl>	pop_million <dbl>
Afghanistan	Asia	1952	28.80100	8425333	779.4453	8.425333
Afghanistan	Asia	1957	30.33200	9240934	820.8530	9.240934
Afghanistan	Asia	1962	31.99700	10267083	853.1007	10.267083
Afghanistan	Asia	1967	34.02000	11537966	836.1971	11.537966
Afghanistan	Asia	1972	36.08800	13079460	739.9811	13.079460
Afghanistan	Asia	1977	38.43800	14880372	786.1134	14.880372
Afghanistan	Asia	1982	39.85400	12881816	978.0114	12.881816
Afghanistan	Asia	1987	40.82200	13867957	852.3959	13.867957
Afghanistan	Asia	1992	41.67400	16317921	649.3414	16.317921
Afghanistan	Asia	1997	41.76300	22227415	635.3414	22.227415

1-10 of 1,704 rows

Previous 1 2 3 4 5 6 ... 100 Next

# MUTATE

```
gapminder %>%  
  mutate(lifeExp_binary = ifelse(lifeExp < 40,  
                                "Very low", "Not very low"))
```

country <fctr>	continent <fctr>	year <int>	lifeExp <dbl>	pop <int>	gdpPercap <dbl>	lifeExp_binary <chr>
Afghanistan	Asia	1952	28.80100	8425333	779.4453	Very low
Afghanistan	Asia	1957	30.33200	9240934	820.8530	Very low
Afghanistan	Asia	1962	31.99700	10267083	853.1007	Very low
Afghanistan	Asia	1967	34.02000	11537966	836.1971	Very low
Afghanistan	Asia	1972	36.08800	13079460	739.9811	Very low
Afghanistan	Asia	1977	38.43800	14880372	786.1134	Very low
Afghanistan	Asia	1982	39.85400	12881816	978.0114	Very low
Afghanistan	Asia	1987	40.82200	13867957	852.3959	Not very low
Afghanistan	Asia	1992	41.67400	16317921	649.3414	Not very low
Afghanistan	Asia	1997	41.76300	22227415	635.3414	Not very low

1-10 of 1,704 rows

Previous 1 2 3 4 5 6 ... 100 Next

# GROUP\_BY + SUMMARIZE

```
gapminder %>%
  group_by(continent) %>%
  summarize(avg_lifeexp = mean(lifeExp),
            median_lifeexmp = median(lifeExp),
            num_countries = n())
```

continent <fctr>	avg_lifeexp <dbl>	median_lifeexmp <dbl>	num_countries <int>
Africa	48.86533	47.7920	624
Americas	64.65874	67.0480	300
Asia	60.06490	61.7915	396
Europe	71.90369	72.2410	360
Oceania	74.32621	73.6650	24

5 rows

# GROUP\_BY + SUMMARIZE

```
gapminder %>%
  group_by(continent, year) %>%
  summarize(avg_lifeexp = mean(lifeExp),
            median_lifeexmp = median(lifeExp),
            num_countries = n())
```

continent <fctr>	year <int>	avg_lifeexp <dbl>	median_lifeexmp <dbl>	num_countries <int>
Africa	1952	39.13550	38.8330	52
Africa	1957	41.26635	40.5925	52
Africa	1962	43.31944	42.6305	52
Africa	1967	45.33454	44.6985	52
Africa	1972	47.45094	47.0315	52
Africa	1977	49.58042	49.2725	52
Africa	1982	51.59287	50.7560	52
Africa	1987	53.34479	51.6395	52
Africa	1992	53.62958	52.4290	52
Africa	1997	53.59827	52.7590	52

# OTHER HELPFUL VERBS

---

`arrange()`

Sort a data frame by a column

`left_join()`

Merge two data frames by column(s)

`count()`

`group_by() %>% summarize(n = n())`

`gather()`

Make a data frame long

`spread()`

Make a data frame wide

# LIVE EXAMPLE