

Data Visualization with R and ggplot

Jan Zilinsky

Table of contents

Overview of this class	4
Lesson 1: Think of output from statistical models as data. Summarize it visually. . .	4
Lesson 2: The first chart you produce is your “draft #1”. Iterate and create reasonably clear and self-contained visual exhibits.	4
Introductory topics	5
A ggplot deep dive	5
Visualizing output from statistical models	5
1 Principles	6
1.1 There are always tradeoffs	6
2 Toplines and crosstabs	13
2.1 Favorability by vote choice (intent)	14
2.2.1 List of variables	16
2.3 We subset the data to the first 10 waves	27
2.9.1 Favorability of AOC	32
2.11 Favorability by vote choice (intent)	34
2.12 Favorability by vote choice (intent)	35
2.13 Favorability by vote choice (intent)	35
2.14 Favorability by vote choice (intent)	35
2.15 Favorability by vote choice (intent)	35
2.16 Favorability by vote choice (intent)	36
2.17 Favorability by vote choice (intent)	37
2.18 Favorability by vote choice (intent)	38
2.19 Favorability by vote choice (intent)	39
2.20 Favorability by vote choice (intent)	40
2.21 Favorability by vote choice (intent)	41
2.22 Favorability by vote choice (intent)	42
3 AJPS 2021 dataset	44
4 Economic data over time	46
4.0.1 Zomm in on OECD countries and increase the time horizon	52
5 Standard charts	56

6	Advanced ggplot	57
7	Visualizing statistical models	58
	References	59
	Useful resources include:	59

Overview of this class

Creating effective visualizations of social and political data can help you discover and communicate new insights. This website contains much of the material I use in my course designed to help students become better communicators with R. Most of what students would need to do can be accomplished with `ggplot2`.

If I had to compress the content of this course into just two lessons, I would say this:

Lesson 1: Think of output from statistical models as data. Summarize it visually.

There is a lot we can do after extracting relevant information from statistical models. What is the distribution of predictions when you change your machine learning algorithm? How stable are coefficients from linear models across specifications. What do we lose/gain when we display aggregates (e.g. average marginal effects) rather than individual data-points?

Lesson 2: The first chart you produce is your “draft #1”. Iterate and create reasonably clear and self-contained visual exhibits.

Students are encouraged to think creatively about visualizing different types of information and data. Significant coding experience with R is expected, but prior knowledge of issues related to design, style, and principles of data visualization is not required.

After taking this course, students will be expected to be able to present real data clearly and to identify strengths and weakness of existing data displays and dashboards.

We will be graphing various types of evidence including:

- public opinion data
- macro-economic data
- summaries of statistical models
- quantitative representations of text (e.g. content of social media post and the accompanying

Introductory topics

- What works and what to avoid even if it works?
- Principles of visual perception and effect communication
- Getting familiar with ggplot

A ggplot deep dive

- Toplines, cross-tabs
- Geometries, statistics and coordinates
- Facets, themes
- Refining plots
- 3-way cross-tabs
- Heatmaps

Visualizing output from statistical models

- Coefficients and uncertainty
- Predicted probabilities, marginal effects, and interactions
- Model performance (in-sample and out-of-sample comparisons)
- Machine learning output (regression trees, most important variables, etc.)

Assignments:

- Create your own dataset (30%).
 - Create your own dataset. It needs to have at least one of these 3 attributes
 1. Multiple levels (at least 2).
 2. Original topic, subject or angle.
 3. Impressive scope
- Final project (70%)
 - Form a group of 2-3 students
 - Prepare a compelling data visualization
 - Some elements in R are expected, you could also use [D3](#) or another language if you wish.

1 Principles

1.1 There are always tradeoffs

A frequent tradeoff can be thought of as **truthfulness** vs. simplicity.

In other words, you will need to balance:

- Readability vs. “completeness”
- Concise vs. “attention-gabbing”
- Simplicity vs. other goals

If you drop outliers, for example, your chart’s readability will almost surely improve. But it could be less truthful.

If you label a subset of your observations, then arguably some information “is lost”, unless you post your data.

But I would still assert that you should almost ever make graphs that look like this:

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr      1.1.1      v readr      2.1.4
v forcats    1.0.0      v stringr    1.5.0
v ggplot2    3.4.2      v tibble     3.2.1
v lubridate  1.9.2      v tidyr      1.3.0
v purrr      1.0.1
```

```
-- Conflicts ----- tidyverse_conflicts() --
```

```
x dplyr::filter() masks stats::filter()
```

```
x dplyr::lag()     masks stats::lag()
```

```
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
```

```
# Get fiscal data
```

```
imf <- read_csv("data_macro/imf-fiscalmonitor-apr2023.csv")
```

Rows: 108 Columns: 12

-- Column specification -----

Delimiter: ","

chr (2): variable, country

dbl (10): 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023

i Use `spec()` to retrieve the full column specification for this data.

i Specify the column types or set `show_col_types = FALSE` to quiet this message.

imf

A tibble: 108 x 12

	variable	country	`2014`	`2015`	`2016`	`2017`	`2018`	`2019`	`2020`	`2021`
	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	General Gove~	Austra~	36.9	37.4	37.4	36.9	37.0	39.1	44.6	42.2
2	General Gove~	Austria	52.3	51.0	50.1	49.3	48.8	48.6	56.7	56.0
3	General Gove~	Belgium	55.6	53.7	53.1	52.0	52.3	51.9	58.9	55.5
4	General Gove~	Canada	38.4	40.0	40.8	40.5	40.7	40.6	52.7	45.9
5	General Gove~	Croatia	48.6	47.5	46.4	44.7	45.5	46.1	54.0	48.5
6	General Gove~	Cyprus	40.3	39.3	37.3	36.4	42.6	38.1	44.6	43.1
7	General Gove~	Czechia	42.6	41.9	39.8	39.0	40.6	41.1	47.2	46.5
8	General Gove~	Denmark	55.2	54.5	52.5	50.5	50.5	49.7	53.5	50.8
9	General Gove~	Estonia	37.8	39.5	39.4	39.2	39.3	39.4	44.8	41.3
10	General Gove~	Finland	57.3	56.5	55.6	53.6	53.4	53.3	57.1	55.7

i 98 more rows

i 2 more variables: `2022` <dbl>, `2023` <dbl>

Reformat data to wide and keep only the latest year

```
imf_wide2022 <- imf %>% select(variable, country, `2022`) %>%  
  pivot_wider(names_from = variable, values_from = `2022`)
```

imf_wide2022

A tibble: 36 x 4

	country	General Government Expendit~1	`Overall Balance`	`Gross public debt`
	<chr>	<dbl>	<dbl>	<dbl>
1	Australia	38.4	-3.30	55.7
2	Austria	52.4	-3.31	77.8
3	Belgium	53.7	-4.32	105.

4	Canada	41.5	-0.700	107.
5	Croatia	45.7	-0.943	67.5
6	Cyprus	39.9	2.26	86.5
7	Czechia	44.8	-3.59	42.3
8	Denmark	49.2	2.48	29.7
9	Estonia	40.2	-1.15	17.2
10	Finland	54.0	-1.86	74.8

i 26 more rows
i abbreviated name: 1: `General Government Expenditure`

```
# Merge in inflation data:
inf <- read_csv("data_macro/inflation_WDI.csv")
```

```
Rows: 8522 Columns: 4
-- Column specification -----
Delimiter: ","
chr (2): country, iso3c
dbl (2): inflation, year

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

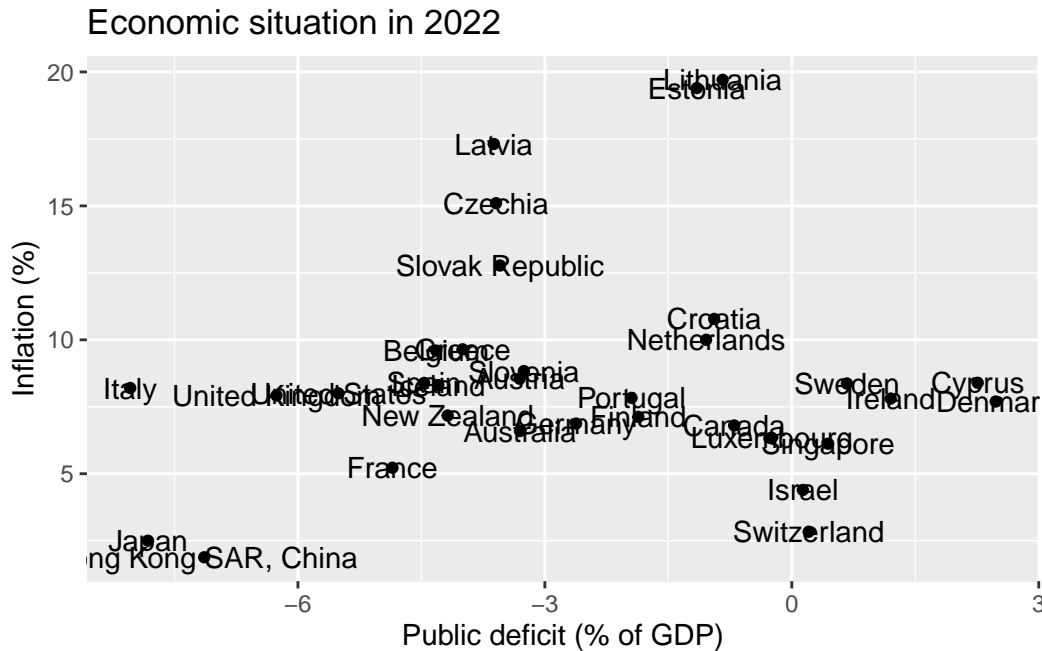
```
inf2022 <- inf %>% filter(year==2022)

econ2022 <- left_join(imf_wide2022,inf2022,by="country")

# Make a simple plot:
econ2022 %>%
  filter(country != "Norway") %>%
  ggplot(aes(y=inflation,
             x=`Overall Balance`,
             label=country)) +
  geom_point() +
  geom_text() +
  labs(title = "Economic situation in 2022",
       x="Public deficit (% of GDP)", y= "Inflation (%)")
```

Warning: Removed 2 rows containing missing values (`geom_point()`).

Warning: Removed 2 rows containing missing values (``geom_text()``).



Starting in 2021, inflation increased in many countries and became a source of serious concern for citizens and politicians. One set of substantive questions dealt with this: should governments be blamed for excessive spending (and borrowing)? Were fiscal decisions responsible for inflation? Here, we'll deal with potential approach to design visual exhibits that might facilitate some international comparisons.^[1] But also remember that these type of evidence can inform factual debates but they wouldn't settle the debate.]

This chart plots the relationship between between public budget deficits and inflation.

A few things to notice:

- We are not displaying Norway (can you check why?)
- We added an informative title to the scatterplot
- We made the questionable choice to use `geom_text()` which uses what we placed inside `aes(... label=country)`.

The same data can be displayed this way:

```
econ2022 %>%  
  filter(country != "Norway") %>%  
  ggplot(aes(y=inflation,  
             x=`Overall Balance`,
```

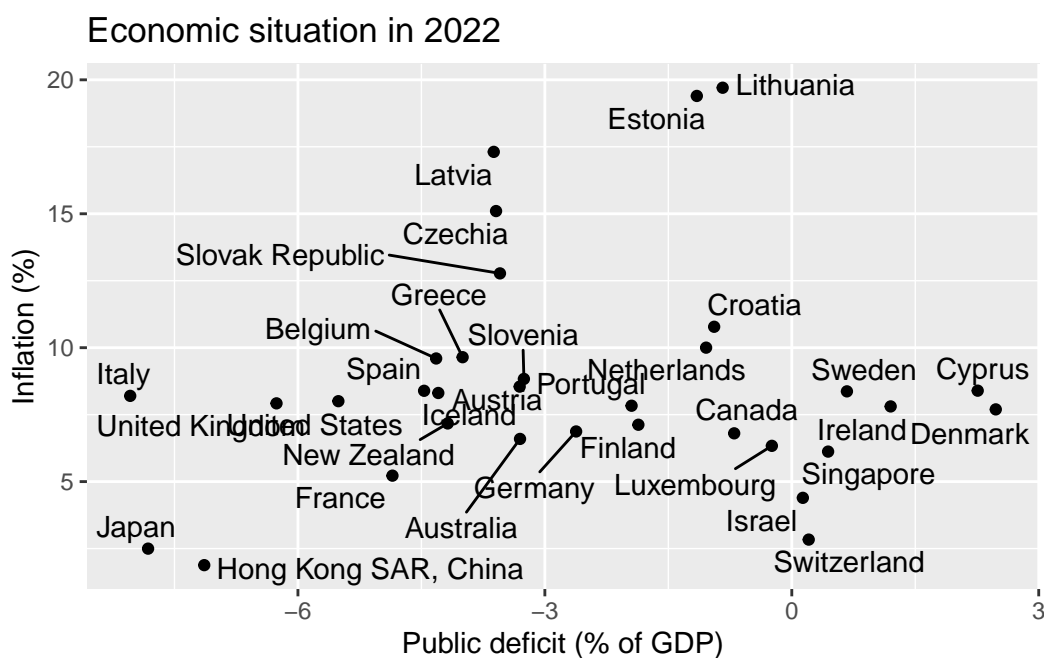
```

    label=country)) +
  geom_point() +
  ggrepel::geom_text_repel() +
  labs(title = "Economic situation in 2022",
       x="Public deficit (% of GDP)", y= "Inflation (%)")

```

Warning: Removed 2 rows containing missing values (`geom_point()`).

Warning: Removed 2 rows containing missing values (`geom_text_repel()`).



Or you can highlight a subset of subset of observations relevant for your analysis:

```

econ2022 %>%
  filter(country != "Norway") %>%
  ggplot(aes(y=inflation,
             x=`Overall Balance`)) +
  geom_point() +
  ggrepel::geom_text_repel(data=econ2022 %>%
                           filter(country %in% c("Italy",
                                                  "Sweden",
                                                  "United States"))),

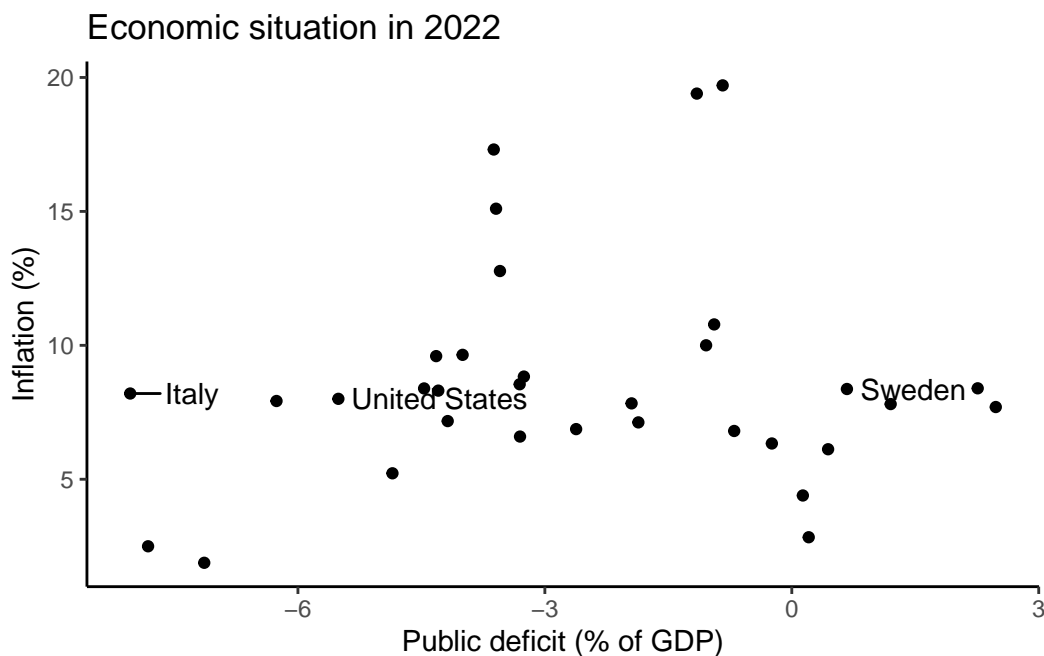
```

```

aes(label=country),nudge_x=.75) +
labs(title = "Economic situation in 2022",
      x="Public deficit (% of GDP)", y= "Inflation (%)") +
theme_classic()

```

Warning: Removed 2 rows containing missing values (`geom_point()`).



We see that deficit spending is not informative: moderately high inflation was common across OECD countries, and neither large deficits, nor budget surpluses, were prognostic of better/worse outcomes.

What about total government spending?

```

econ2022 %>%
  ggplot(aes(y=inflation,
             x=`General Government Expenditure`)) +
  geom_point() +
  ggrepel::geom_text_repel(data=econ2022 %>%
                           filter(country %in% c("Italy",
                                                  "Sweden",
                                                  "United States"))),

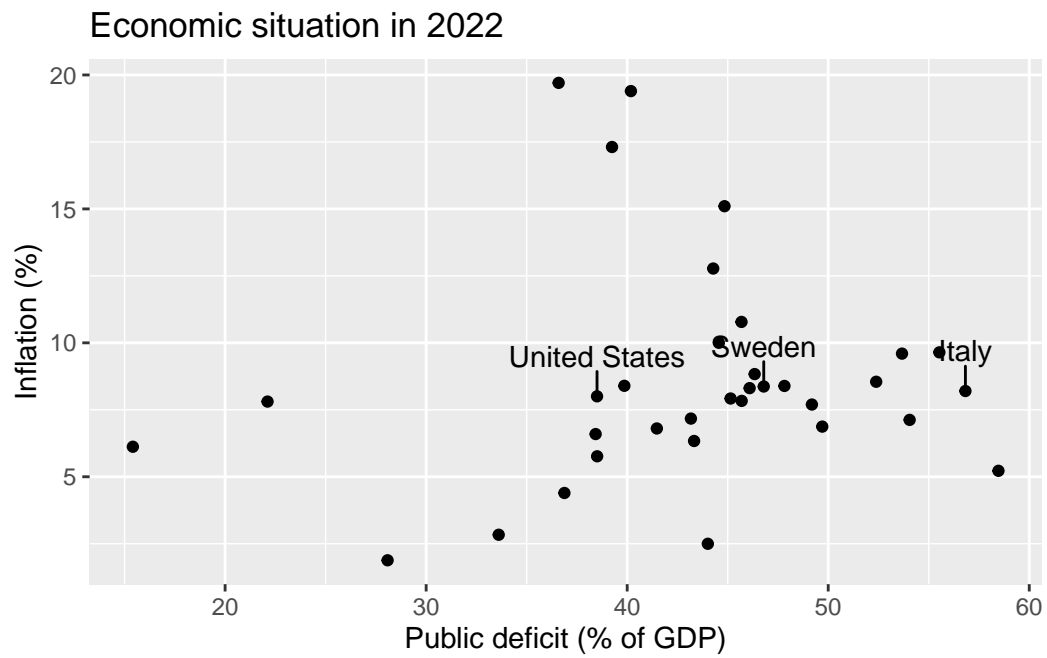
```

```

aes(label=country),nudge_y=1.5) +
labs(title = "Economic situation in 2022",
      x="Public deficit (% of GDP)", y= "Inflation (%)")

```

Warning: Removed 2 rows containing missing values (`geom_point()`).



2 Toplines and crosstabs

In many, maybe most, situations, your dataset will have so many rows that it would not be realistic to plot every single observations.

This means that you will want to:

1. Calculate some aggregate statistics first. Think of this “new” data object as just a new spreadsheet.
2. Pass the smaller object (often a tibble or a data frame) into `ggplot()`. This the first layer.
3. Choose the parts (rows and columns) which are most relevant.
4. Experiment with...
 - ... arrangements within `aes()`
 - ... different geoms
 - ... sub-groups of the data (faceting may or may not be informative)
 - ... themes, labels, and so on.

```
library(tidyverse)
```

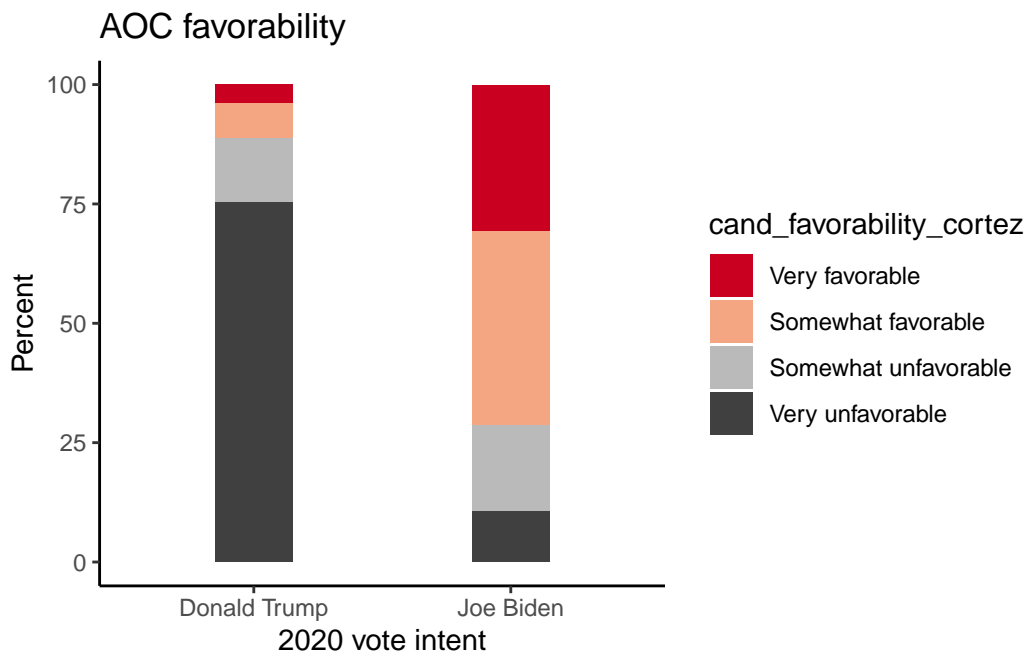
```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr      1.1.1      v readr      2.1.4
v forcats    1.0.0      v stringr    1.5.0
v ggplot2    3.4.2      v tibble     3.2.1
v lubridate  1.9.2      v tidyr      1.3.0
v purrr      1.0.1
-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()     masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
```

```
library(haven)
library(labelled)
library(pollster)
```

```
a <- readRDS("data_nationscape2019/Nationscape_first10waves.rds")
```

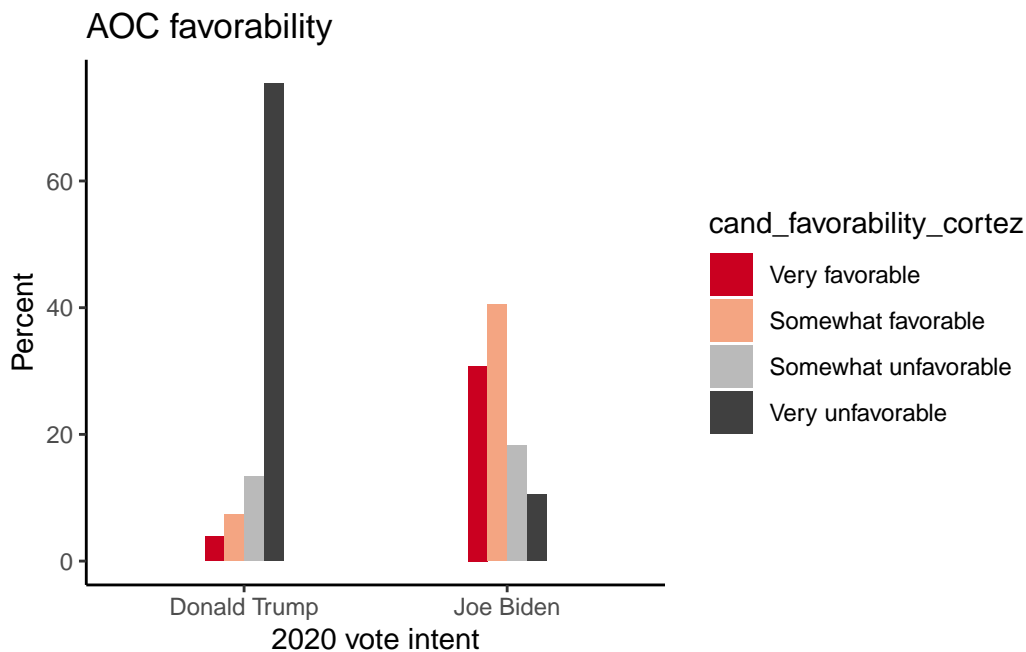
2.1 Favorability by vote choice (intent)

```
pollster::crosstab(df = a,  
  x = trump_biden,  
  y = cand_favorability_cortez,  
  weight = weight,  
  format = "long") %>%  
filter(trump_biden!="Don't Know") %>%  
ggplot(aes(x= fct_reorder(trump_biden,pct),  
  y = pct,fill=cand_favorability_cortez)) +  
geom_col(width = .3) +  
  theme_classic() +  
scale_fill_brewer(palette = "RdGy") +  
labs(y="Percent", x = "2020 vote intent", title = "AOC favorability")
```

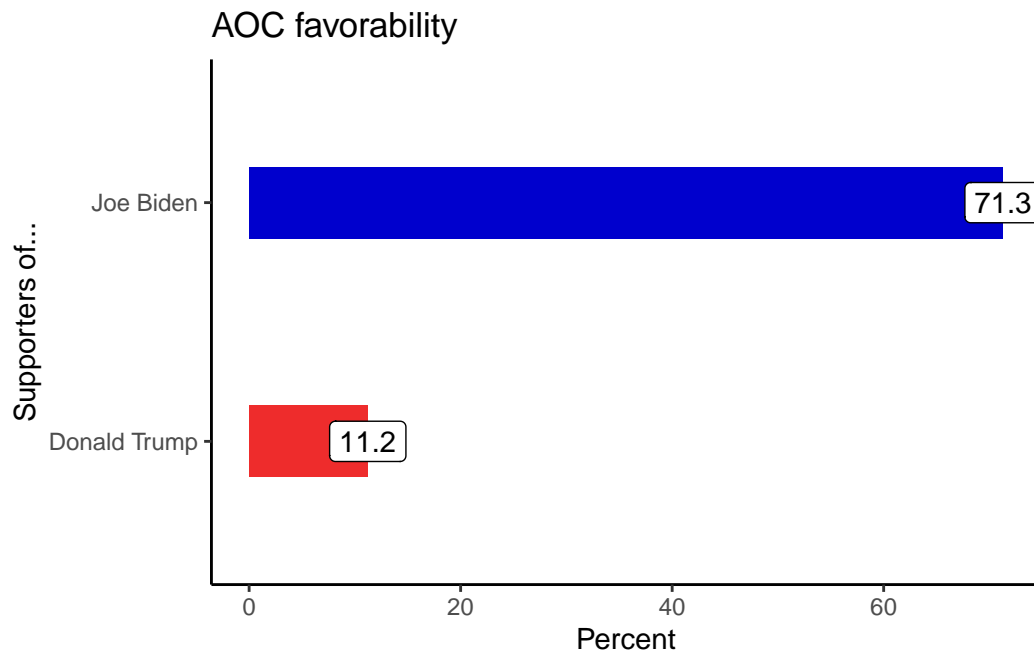


```
pollster::crosstab(df = a,  
  x = trump_biden,  
  y = cand_favorability_cortez,  
  weight = weight,  
  format = "long") %>%  
filter(trump_biden!="Don't Know") %>%
```

```
ggplot(aes(x= fct_reorder(trump_biden,pct),
                    y = pct,fill=cand_favorability_cortez)) +
  geom_col(width = .3, position = position_dodge()) +
  theme_classic() +
  scale_fill_brewer(palette = "RdGy") +
  labs(y="Percent", x = "2020 vote intent", title = "AOC favorability")
```



```
pollster::crosstab(df = a,
                   x = trump_biden,
                   y = aoc_Favorable,
                   weight = weight,
                   format = "long") %>%
  filter(aoc_Favorable==1, trump_biden!="Don't Know") %>%
  ggplot(aes(y= fct_reorder(trump_biden,pct),
                    x = pct)) +
  geom_col(width = .3,fill=c("blue3","firebrick2")) +
  theme_classic() +
  labs(x="Percent", y = "Supporters of...", title = "AOC favorability") +
  geom_label(aes(label=round(pct,1)))
```



2.2

2.2.1 List of variables

```
dim(a)
```

```
[1] 62408 889
```

```
names(a)
```

```
[1] "response_id"           "start_date"
[3] "right_track"           "economy_better"
[5] "interest"              "registration"
[7] "news_sources_facebook" "news_sources_cnn"
[9] "news_sources_msnbc"    "news_sources_fox"
[11] "news_sources_network"  "news_sources_localtv"
[13] "news_sources_telemundo" "news_sources_npr"
[15] "news_sources_amtalk"   "news_sources_new_york_times"
[17] "news_sources_local_newspaper" "news_sources_other"
```


[19]	"news_sources_other_TEXT"	"pres_approval"
[21]	"vote_intention"	"vote_2016"
[23]	"vote_2016_other_text"	"consider_trump"
[25]	"not_trump"	"primary_party"
[27]	"group_favorability_whites"	"group_favorability_blacks"
[29]	"group_favorability_latinos"	"group_favorability_asians"
[31]	"group_favorability_christians"	"group_favorability_socialists"
[33]	"group_favorability_muslims"	"group_favorability_labor_unions"
[35]	"group_favorability_the_police"	"group_favorability_undocumented"
[37]	"group_favorability_lgbt"	"group_favorability_republicans"
[39]	"group_favorability_democrats"	"cand_favorability_trump"
[41]	"cand_favorability_obama"	"cand_favorability_cortez"
[43]	"cand_favorability_biden"	"cand_favorability_harris"
[45]	"cand_favorability_buttigieg"	"cand_favorability_warren"
[47]	"cand_favorability_sanders"	"cand_favorability_pence"
[49]	"dem_vote_intent"	"dem_vote_intent_TEXT"
[51]	"rank_dems_1"	"rank_dems_2"
[53]	"rank_dems_3"	"replace_trump"
[55]	"house_intent"	"senate_intent"
[57]	"governor_intent"	"trump_biden"
[59]	"trump_sanders"	"trump_harris"
[61]	"trump_warren"	"trump_buttigieg"
[63]	"trump_booker"	"trump_castro"
[65]	"trump_gabbard"	"trump_gillibrand"
[67]	"trump_orourke"	"pence_biden"
[69]	"pence_buttigieg"	"pence_harris"
[71]	"pence_sanders"	"pence_warren"
[73]	"cand_truth_donald_trump"	"cand_truth_elizabeth_warren"
[75]	"cand_truth_joe_biden"	"cand_truth_bernie_sanders"
[77]	"cand_truth_pete_buttigieg"	"cand_truth_kamala_harris"
[79]	"cand_facts_donald_trump"	"cand_facts_elizabeth_warren"
[81]	"cand_facts_joe_biden"	"cand_facts_bernie_sanders"
[83]	"cand_facts_pete_buttigieg"	"cand_facts_kamala_harris"
[85]	"racial_attitudes_tryhard"	"racial_attitudes_generations"
[87]	"racial_attitudes_marry"	"racial_attitudes_date"
[89]	"gender_attitudes_maleboss"	"gender_attitudes_logical"
[91]	"gender_attitudes_opportunity"	"gender_attitudes_complain"
[93]	"discrimination_blacks"	"discrimination_whites"
[95]	"discrimination_muslims"	"discrimination_christians"
[97]	"discrimination_women"	"discrimination_men"
[99]	"sen_knowledge"	"sc_knowledge"
[101]	"pid3"	"pid7_legacy"
[103]	"strength_democrat"	"strength_republican"

[105]	"lean_independent"	"ideo5"
[107]	"employment"	"employment_other_text"
[109]	"foreign_born"	"language"
[111]	"religion"	"religion_other_text"
[113]	"is_evangelical"	"orientation_group"
[115]	"in_union"	"household_gun_owner"
[117]	"wall"	"cap_carbon"
[119]	"environment"	"guns_bg"
[121]	"mctaxes"	"estate_tax"
[123]	"raise_upper_tax"	"college"
[125]	"abortion_waiting"	"abortion_never"
[127]	"abortion_conditions"	"late_term_abortion"
[129]	"abortion_insurance"	"guaranteed_jobs"
[131]	"green_new_deal"	"gun_registry"
[133]	"immigration_separation"	"immigration_system"
[135]	"immigration_wire"	"impeach_trump"
[137]	"israel"	"marijuana"
[139]	"maternityleave"	"medicare_for_all"
[141]	"military_size"	"minwage"
[143]	"muslimban"	"oil_and_gas"
[145]	"reparations"	"right_to_work"
[147]	"ten_commandments"	"trade"
[149]	"trans_military"	"uctaxes2"
[151]	"vouchers"	"gov_insurance"
[153]	"public_option"	"health_subsidies"
[155]	"path_to_citizenship"	"dreamers"
[157]	"deportation"	"ban_guns"
[159]	"ban_assault_rifles"	"limit_magazines"
[161]	"age"	"gender"
[163]	"census_region"	"hispanic"
[165]	"race_ethnicity"	"household_income"
[167]	"education"	"state"
[169]	"congress_district"	"weight"
[171]	"wave"	"extra_ban_discr_race"
[173]	"extra_ban_discr_gend"	"extra_ban_discr_age"
[175]	"extra_women_juries"	"extra_driver_insure"
[177]	"extra_helmet_moto"	"extra_tax_cigarettes"
[179]	"extra_school_silence"	"extra_ban_corp_pun"
[181]	"rep_vote_prim"	"rep_vote_prim_TEXT"
[183]	"group_favorability_evangelicals"	"group_favorability_white_men"
[185]	"fc_smallgov"	"fc_trad_val"
[187]	"statements_protect_traditions"	"statements_defense_burden"
[189]	"statements_trade_effects"	"statements_christianity_assault"

[191]	"statements_gender_identity"	"statements_american_loss"
[193]	"statements_imm_assimilate"	"statements_gun_rights"
[195]	"statements_confront_china"	"statements_foreign_interests"
[197]	"abortion_any_time"	"abolish_priv_insurance"
[199]	"china_tariffs"	"criminal_immigration"
[201]	"immigration_insurance"	"extra_group_favor_democrats"
[203]	"extra_group_favor_republicans"	"primary_sen_barrasso"
[205]	"primary_sen_blackburn"	"primary_sen_blunt"
[207]	"primary_sen_cassidy"	"primary_sen_collins"
[209]	"primary_sen_cornyn"	"primary_sen_cotton"
[211]	"primary_sen_daines"	"primary_sen_ernst"
[213]	"primary_sen_gardner"	"primary_sen_graham"
[215]	"primary_sen_hoeven"	"primary_sen_hydesmith"
[217]	"primary_sen_inhofe"	"primary_sen_lee"
[219]	"primary_sen_mcconnell"	"primary_sen_mcsally"
[221]	"primary_sen_moorecapito"	"primary_sen_moran"
[223]	"primary_sen_perdue"	"primary_sen_portman"
[225]	"primary_sen_risch"	"primary_sen_rounds"
[227]	"primary_sen_rubio"	"primary_sen_sasse"
[229]	"primary_sen_shelby"	"primary_sen_sullivan"
[231]	"primary_sen_tillis"	"primary_sen_toomey"
[233]	"primary_sen_young"	"primary_sen_boozman"
[235]	"primary_sen_braun"	"primary_sen_cramer"
[237]	"primary_sen_crapo"	"primary_sen_cruz"
[239]	"primary_sen_fischer"	"primary_sen_grassley"
[241]	"primary_sen_hawley"	"primary_sen_lankford"
[243]	"primary_sen_murkowski"	"primary_sen_neelykennedy"
[245]	"primary_sen_paul"	"primary_sen_romney"
[247]	"primary_sen_scott_rick"	"primary_sen_scott_tim"
[249]	"primary_sen_thune"	"primary_sen_wicker"
[251]	"extra_group_favor_demcong"	"extra_group_favor_repcong"
[253]	"extra_group_identity_race"	"extra_group_identity_partyID"
[255]	"extra_group_id_relig_belief"	"extra_group_identity_local_comm"
[257]	"extra_group_identity_close_fam"	"extra_group_id_being_american"
[259]	"extra_n_children"	"extra_n_adults"
[261]	"extra_housing"	"extra_party_ideo_rate_dem"
[263]	"extra_party_ideo_rate_rep"	"extra_party_ideo_dem_desire"
[265]	"extra_party_ideo_rep_desire"	"extra_foreign_interf"
[267]	"extra_foreign_help"	"extra_foreign_effect"
[269]	"vote_2012"	"vote_2012_other_text"
[271]	"extra_dem_prim"	"extra_rep_prim"
[273]	"extra_dem_prim_cand"	"extra_rep_prim_cand"
[275]	"extra_cong_race_white"	"extra_cong_race_black_or_africa"

[277]	"extra_cong_race_hispanic_or_lat"	"extra_cong_race_asian"
[279]	"extra_cong_party_democrats"	"extra_cong_party_republicans"
[281]	"extra_county_imm"	"extra_neighb_race_white"
[283]	"extra_neighb_race_black_or_afri"	"extra_neighb_race_hispanic_or_l"
[285]	"extra_neighb_race_asian"	"extra_neighb_party_democrats"
[287]	"extra_neighb_party_republicans"	"extra_neighb_imm"
[289]	"extra_trump_cares_me"	"extra_trump_cares_poor"
[291]	"extra_trump_cares_mid_class"	"extra_trump_cares_wealthy"
[293]	"extra_biden_cares_me"	"extra_biden_cares_poor"
[295]	"extra_biden_cares_mid_class"	"extra_biden_cares_wealthy"
[297]	"extra_warren_cares_me"	"extra_warren_cares_poor"
[299]	"extra_warren_cares_mid_class"	"extra_warren_cares_wealthy"
[301]	"extra_contact_dem"	"extra_contact_rep"
[303]	"extra_contact_third"	"extra_contact_nonpart"
[305]	"extra_cont_dem_mode_in_person"	"extra_cont_dem_mode_phone_call"
[307]	"extra_cont_dem_mode_email_or_te"	"extra_cont_dem_mode_letter_or_p"
[309]	"extra_cont_rep_mode_in_person"	"extra_cont_rep_mode_phone_call"
[311]	"extra_cont_rep_mode_email_or_te"	"extra_cont_rep_mode_letter_or_p"
[313]	"extra_cont_third_mode_in_person"	"extra_cont_third_mode_phone_cal"
[315]	"extra_cont_third_mode_email_or_"	"extra_cont_third_mode_letter_or"
[317]	"extra_cont_nopart_mode_in_perso"	"extra_cont_nopart_mode_phone_ca"
[319]	"extra_cont_nopart_mode_email_or"	"extra_cont_nopart_mode_letter_o"
[321]	"group_favorability_jews"	"discrimination_jews"
[323]	"saudi_arabia"	"egypt"
[325]	"pid7"	"extra_priorities_deficit"
[327]	"extra_priorities_crime"	"extra_priorities_economy"
[329]	"extra_priorities_immigration"	"extra_priorities_healthcare"
[331]	"extra_deficit_change"	"extra_deficit_raise_taxes"
[333]	"extra_deficit_cut_military"	"extra_deficit_cut_medicaid"
[335]	"extra_deficit_cut_medicare"	"extra_voting_right"
[337]	"extra_vote_fine"	"extra_civic_duty"
[339]	"extra_gov_rep"	"extra_gov_confidence"
[341]	"extra_combat_vote_suppression"	"extra_win_elections"
[343]	"extra_right_notvote"	"extra_uneducated_vote"
[345]	"extra_punish_vote"	"extra_lose_elections"
[347]	"extra_excessive_gov_fines"	"cand_favorability_bloomberg"
[349]	"trump_bloomberg"	"extra_trump_economy"
[351]	"extra_trump_health"	"extra_trump_foreign"
[353]	"extra_trump_immigration"	"extra_trump_trade"
[355]	"extra_trump_race"	"extra_trump_taxes"
[357]	"extra_trump_climate"	"extra_trump_budget"
[359]	"primary_party_past"	"dem_vote_alreadyvoted"
[361]	"dem_vote_already_other_text"	"rep_vote_prim_prospect"

[363]	"extra_race_white_you"	"extra_race_white_mom"
[365]	"extra_race_white_dad"	"extra_race_black_you"
[367]	"extra_race_black_mom"	"extra_race_black_dad"
[369]	"extra_race_latino_you"	"extra_race_latino_mom"
[371]	"extra_race_latino_dad"	"extra_race_asian_you"
[373]	"extra_race_asian_mom"	"extra_race_asian_dad"
[375]	"extra_race_am_indian_you"	"extra_race_am_indian_mom"
[377]	"extra_race_am_indian_dad"	"extra_race_pi_you"
[379]	"extra_race_pi_mom"	"extra_race_pi_dad"
[381]	"extra_race_middle_eastern_you"	"extra_race_middle_eastern_mom"
[383]	"extra_race_middle_eastern_dad"	"extra_race_other_you"
[385]	"extra_race_other_mom"	"extra_race_other_dad"
[387]	"extra_race_dk_you"	"extra_race_dk_mom"
[389]	"extra_race_dk_dad"	"extra_ideo_trump"
[391]	"extra_ideo_warren"	"extra_ideo_buttigieg"
[393]	"extra_ideo_sanders"	"extra_ideo_bloomberg"
[395]	"extra_ideo_klobuchar"	"extra_ideo_biden"
[397]	"extra_pop_latinos"	"extra_pop_women"
[399]	"extra_pop_young"	"extra_pop_afam"
[401]	"extra_pop_men"	"extra_sanders_cares_me"
[403]	"extra_sanders_cares_poor"	"extra_sanders_cares_mid_class"
[405]	"extra_sanders_cares_wealthy"	"extra_bloomb_cares_me"
[407]	"extra_bloomb_cares_poor"	"extra_bloomb_cares_mid_class"
[409]	"extra_bloomb_cares_wealthy"	"dem_vote_intent_other_text"
[411]	"cand_favorability_klobuchar"	"cand_favorability_abrams"
[413]	"extra_prescriptions"	"extra_covid_cancel_meet"
[415]	"extra_covid_close_business"	"extra_covid_close_schools"
[417]	"extra_covid_work_home"	"extra_covid_restrict_travel"
[419]	"extra_covid_restrict_home"	"extra_covid_social_distance"
[421]	"extra_covid_delay_elections"	"extra_covid_testing"
[423]	"extra_covid_gov_travel"	"extra_covid_increase_gov_unemp"
[425]	"extra_covid_gov_health"	"extra_covid_gov_sick_leave"
[427]	"extra_covid_gov_interest"	"extra_covid_gov_cash"
[429]	"extra_corona_concern"	"extra_sick_you"
[431]	"extra_sick_family"	"extra_sick_work"
[433]	"extra_sick_other"	"extra_covid_wash"
[435]	"extra_covid_cancel_travel"	"extra_covid_stock_goods"
[437]	"extra_covid_visit_family"	"extra_covid_quarantine"
[439]	"extra_covid_hospital"	"extra_income_corona"
[441]	"extra_trump_corona"	"dem_vote_past"
[443]	"dem_vote_past_other_text"	"rep_vote_past"
[445]	"rep_vote_past_TEXT"	"extra_worry_finances"
[447]	"extra_savings"	"extra_prepare_money"

[449]	"extra_lost_job"	"extra_spouse_lost_job"
[451]	"extra_mortgage_difficulty"	"extra_stud_loan_difficulty"
[453]	"extra_car_pay_difficulty"	"extra_drop_income"
[455]	"extra_satisf_job"	"extra_satisf_income"
[457]	"extra_satisf_savings"	"extra_satisf_debt"
[459]	"discrimination_asians"	"extra_rumor_conceal"
[461]	"extra_rumor_lab"	"extra_rumor_politics"
[463]	"extra_rumor_vaccine"	"extra_rumor_treatment"
[465]	"extra_rumor_infect_under30"	"extra_rumor_safe_under30"
[467]	"extra_gub_corona"	"extra_local_corona"
[469]	"extra_covid_airlines"	"extra_covid_small_business"
[471]	"pence_bloomberg"	"extra_tolerance_econ"
[473]	"extra_return_restaurant"	"extra_return_sports"
[475]	"extra_return_child_school"	"extra_return_concert"
[477]	"extra_return_fly"	"extra_return_shop"
[479]	"extra_return_public_transit"	"extra_return_movies"
[481]	"extra_return_dinner"	"extra_return_dentist"
[483]	"extra_return_haircut"	"extra_return_funeral"
[485]	"extra_return_wedding"	"extra_return_church"
[487]	"extra_return_school"	"extra_corona_deaths"
[489]	"extra_corona_lift"	"extra_corona_relig"
[491]	"extra_timeline_business"	"extra_timeline_schools"
[493]	"extra_timeline_gatherings"	"extra_timeline_stayhome"
[495]	"extra_timeline_elections"	"extra_covid_wear_mask"
[497]	"extra_covid_worn_mask"	"extra_covid_socialize_distance"
[499]	"extra_covid_socialize_no_dist"	"extra_covid_nonessential_goods"
[501]	"vote_2020_v1"	"vote_2020_other_text"
[503]	"rep_prim_vote"	"rep_prim_vote_TEXT"
[505]	"dem_prim_vote"	"dem_prim_vote_TEXT"
[507]	"extra_dem_violence"	"extra_ind_violence"
[509]	"extra_rep_violence"	"extra_rep_victory"
[511]	"extra_ind_victory"	"extra_dem_victory"
[513]	"extra_govern_leader"	"extra_govern_army"
[515]	"extra_govern_democracy"	"extra_protest"
[517]	"vote_2020"	"vote_2020_lean"
[519]	"group_favorability_blm"	"elect_conf_conduct"
[521]	"elect_conf_vote"	"consider_biden"
[523]	"not_biden"	"extra_corona_cong"
[525]	"cand_favorability_rice"	"cand_favorability_bass"
[527]	"extra_credit_card"	"extra_vote_ever"
[529]	"extra_vote_mail_past"	"extra_vote_mail_2020"
[531]	"extra_ideo_rate_trump"	"extra_ideo_rate_biden"
[533]	"extra_ideo_rate_harris"	"extra_ideo_rate_pence"

[535]	"extra_dnc_watch"	"extra_rnc_watch"
[537]	"extra_usps_impact"	"extra_vacc_flu"
[539]	"extra_vacc_covid"	"extra_vacc_reason_safety"
[541]	"extra_vacc_reason_never"	"extra_vacc_reason_hadit"
[543]	"extra_vacc_reason_cost"	"extra_vacc_reason_dontcare"
[545]	"extra_vacc_reason_sick"	"extra_vacc_reason_time"
[547]	"extra_vacc_reason_noreason"	"extra_vacc_reason_other"
[549]	"extra_vacc_reason_wontwork"	"extra_vacc_reason_immune"
[551]	"extra_vacc_reason_willget"	"extra_vacc_reason_other_text"
[553]	"extra_age_children_0_5yr"	"extra_age_children_6_10yr"
[555]	"extra_age_children_11_13yr"	"extra_age_children_14_18yr"
[557]	"extra_biden_cares_covid"	"extra_biden_cares_jobs"
[559]	"extra_trump_cares_covid"	"extra_trump_cares_jobs"
[561]	"extra_vote_mail_retr"	"vote_2020_clinton"
[563]	"vote_2020_clinton_other_text"	"vote_2020_clinton_ln"
[565]	"extra_trump_crime"	"vote_2020_retro"
[567]	"vote_2020_retro_other_text"	"extra_covid_require_mask"
[569]	"covid_support_mask"	"covid_pres_mask"
[571]	"covid_support_close"	"covid_pres_close"
[573]	"elect_loser_concede"	"vote_intention_retro"
[575]	"who_won"	"who_won_other_text"
[577]	"primary_party_retro"	"group_favorability_trump_s"
[579]	"group_favorability_biden_s"	"house_intent_retro"
[581]	"senate_intent_retro"	"governor_intent_retro"
[583]	"ga_runoff_intent"	"ga_special_intent"
[585]	"discrimination_latinos"	"married"
[587]	"elect_conf_conduct_retro"	"elect_conf_vote_retro"
[589]	"elect_exp_id"	"elect_exp_polling"
[591]	"elect_exp_reg_deadline"	"elect_exp_harassed"
[593]	"elect_exp_name_reg"	"elect_exp_no_mail"
[595]	"elect_exp_line"	"switchers_text_biden"
[597]	"switchers_text_trump"	"weight_2020"
[599]	"weight_both"	"work_location"
[601]	"pol_rep_minwage"	"pol_rep_guns_bg"
[603]	"pol_rep_uctaxes2"	"pol_rep_public_option"
[605]	"pol_rep_environment"	"pol_rep_ten_commandments"
[607]	"pol_rep_abortion_waiting"	"pol_rep_immigration_wire"
[609]	"pol_rep_estate_tax"	"pol_rep_right_to_work"
[611]	"pol_dem_minwage"	"pol_dem_guns_bg"
[613]	"pol_dem_uctaxes2"	"pol_dem_public_option"
[615]	"pol_dem_environment"	"pol_dem_ten_commandments"
[617]	"pol_dem_abortion_waiting"	"pol_dem_immigration_wire"
[619]	"pol_dem_estate_tax"	"pol_dem_right_to_work"

[621]	"pence_president"	"capitol_approval"
[623]	"capitol_trump_approv"	"capitol_trump_more"
[625]	"twitter_ban"	"Jabortion_latepermit_DISAGREE"
[627]	"Jabortion_never_agree"	"Jabortion_mostlypermit_DISAGREE"
[629]	"Jguns_bg_DISAGREE"	"Jbanguns_DISAGREE"
[631]	"Jban_AR_DISAGREE"	"Jlimit_magazines_DISAGREE"
[633]	"Jabortion_waiting_agree"	"Jabortion_anytime_DISAGREE"
[635]	"Jgun_registry_DISAGREE"	"Jabortion_insdecline_agree"
[637]	"Jten_commandments_agree"	"Jtrans_military_DISAGREE"
[639]	"Jvouchers_agree"	"culture_sum_major"
[641]	"culture_full_theta"	"culture_factor"
[643]	"Jgov_ins_agree"	"Jpublic_option_agree"
[645]	"Jhealth_subsidies_agree"	"Jabolish_priv_insurance"
[647]	"Jraise_upper_tax_agree"	"Jmctaxes_agree"
[649]	"Jestate_NOT_repeal"	"Jcollege_agree"
[651]	"Jcap_agree"	"Jgreen_tech_agree"
[653]	"Jguaranteedjob_agree"	"Jgreen_new_deal_agree"
[655]	"Jmaternityleave_agree"	"Jmedicare_agree"
[657]	"Jminwage_agree"	"Joil_drill_dereg_DISAGREE"
[659]	"Jreparations_agree"	"Jright2W_DISAGREE"
[661]	"Jtrade_limit_agree"	"Juctaxes2_agree"
[663]	"Jchina_tariffs_agree"	"econ_sum_major"
[665]	"z2econMajor8"	"econ_full_theta"
[667]	"econ_cons_theta"	"Jdreamers_path_DISAGREE"
[669]	"Jpath4all_DISAGREE"	"Jwall_agree"
[671]	"Jdeportation_agree"	"Jimmigration_sep_agree"
[673]	"Jimmigration_system_agree"	"Jimmigration_wire_agree"
[675]	"Jmuslimban_agree"	"Jimmigration_insurance_DISAGREE"
[677]	"Jimmigration_felony_agree"	"immig_full_theta"
[679]	"immig_sum_major"	"z2immig4"
[681]	"Jisrael_agree"	"Jmarijuana_agree"
[683]	"Jmilitarysize_agree"	"Jimpeach_agree"
[685]	"cov_airlines_support"	"cov_govcash_support"
[687]	"cov_spendingpoor_support"	"cov_unemployment_support"
[689]	"cov_reduceinterest_support"	"cov_localbusiness_support"
[691]	"covid6_scale"	"covid6_liberalism"
[693]	"z2covid6_liberalism"	"covid6_liberalism_by1SD"
[695]	"racial_attitudes_tryhard_rev"	"racial_attitudes_marry_rev"
[697]	"racial4_liberalism"	"racial4_C"
[699]	"z2racial4_C"	"gender_attitudes_maleboss_rev"
[701]	"gender_attitudes_complain_rev"	"gender_att3"
[703]	"sexism_att3"	"gender_REC_women_complain"
[705]	"gender_REC_woman_boss"	"gender_REC_opportunity_nogood"

[707]	"gender_REC_logcal_not"	"sc_knowledge_correct"
[709]	"news_consumption"	"news_consumption_count"
[711]	"newsconsumption_factor"	"age_group"
[713]	"Men"	"White"
[715]	"Black"	"Asian"
[717]	"Hispanic"	"college_grad"
[719]	"white_noncollege"	"educ_category"
[721]	"inc_group"	"Democrat"
[723]	"Republican"	"emp_fulltime"
[725]	"ADV_nojob"	"ADV_disabled"
[727]	"ADV_any"	"trade_positive_disagree"
[729]	"trump16"	"clinton16"
[731]	"novote2016"	"voted2016"
[733]	"obama12"	"romney12"
[735]	"vote_2016_rec"	"registered"
[737]	"intendsToVote"	"ineligibleToVote"
[739]	"eligibleToVote"	"biden20"
[741]	"warren20"	"sanders20"
[743]	"buttigieg20"	"harris20"
[745]	"wouldvoteforany_BWS"	"trump_v_biden"
[747]	"biden2Pvote_intent"	"trump2Pvote_intent"
[749]	"warren2Pvote_intent"	"sanders2Pvote_intent"
[751]	"buttigieg2Pvote_intent"	"bloomberg2Pvote_intent"
[753]	"vote2020cat"	"trump16toBiden2020"
[755]	"dem_perceived_very_lib"	"rep_perceived_very_cons"
[757]	"trump_strogapprove"	"trump_approve"
[759]	"obama_favorabe"	"biden_favorabe"
[761]	"harris_favorabe"	"warren_favorabe"
[763]	"aoc_Favorable"	"trump_consider"
[765]	"econ_better"	"econ_betterORsame"
[767]	"votcat"	"vot1"
[769]	"vot2"	"vot3"
[771]	"dem_trump16"	"dem_novote2016"
[773]	"dem_clinton2016"	"dem_pref_biden2020"
[775]	"dem_pref_pete2020"	"dem_pref_sanders2020"
[777]	"dem_pref_warren2020"	"dem_pref_harris2020"
[779]	"dem_pref_other2020"	"support_warren_or_sanders"
[781]	"dem_biden2020_top2"	"econ_cons_theta_by1SD"
[783]	"econ_full_theta_by1SD"	"culture_full_theta_by1SD"
[785]	"immig_full_theta_by1SD"	"racial4_C_by1SD"
[787]	"gender_att3_by1SD"	"zt1_econ_full_theta"
[789]	"zt1_culture_full_theta"	"zt1_immig_full_theta"
[791]	"zt1_econ_cons_theta"	"zt1_racial4_C"

[793]	"zt1_gender_att3"	"zt1_sexism_att3"
[795]	"asians_favorable"	"blacks_favorable"
[797]	"whites_favorable"	"democrats_favorable"
[799]	"labor_unions_favorable"	"latinos_favorable"
[801]	"lgbt_favorable"	"muslims_favorable"
[803]	"republicans_favorable"	"socialists_favorable"
[805]	"the_police_favorable"	"undocumented_favorable"
[807]	"white_men_favorable"	"christians_favorable"
[809]	"evangelicals_favorable"	"jews_favorable"
[811]	"asians_favor_withDK"	"blacks_favor_withDK"
[813]	"whites_favor_withDK"	"democrats_favor_withDK"
[815]	"labor_unions_favor_withDK"	"latinos_favor_withDK"
[817]	"lgbt_favor_withDK"	"muslims_favor_withDK"
[819]	"republicans_favor_withDK"	"socialists_favor_withDK"
[821]	"the_police_favor_withDK"	"undocumented_favor_withDK"
[823]	"white_men_favor_withDK"	"christians_favor_withDK"
[825]	"evangelicals_favor_withDK"	"jews_favor_withDK"
[827]	"asians_unfavor_withDK"	"blacks_unfavor_withDK"
[829]	"whites_unfavor_withDK"	"democrats_unfavor_withDK"
[831]	"labor_unions_unfavor_withDK"	"latinos_unfavor_withDK"
[833]	"lgbt_unfavor_withDK"	"muslims_unfavor_withDK"
[835]	"republicans_unfavor_withDK"	"socialists_unfavor_withDK"
[837]	"the_police_unfavor_withDK"	"undocumented_unfavor_withDK"
[839]	"white_men_unfavor_withDK"	"christians_unfavor_withDK"
[841]	"evangelicals_unfavor_withDK"	"jews_unfavor_withDK"
[843]	"s1"	"s2"
[845]	"s3"	"cd"
[847]	"Democrat_w_leaners"	"Republican_w_leaners"
[849]	"pid7_0_to_1"	"date"
[851]	"monthyear"	"week"
[853]	"MY"	"quarter"
[855]	"trump2P_2020retro"	"trump2P_2020plan"
[857]	"approval_trump_scale"	"favorabilityDems"
[859]	"favorabilityReps"	"affPol"
[861]	"unemployed"	"evangelical"
[863]	"fav_blm"	"unfav_blm"
[865]	"unfav_dems"	"unfav_reps"
[867]	"fav_dems"	"fav_reps"
[869]	"fav_Muslims"	"unfav_Muslims"
[871]	"unfav_Asians"	"unfav_Blacks"
[873]	"unfav_Latinos"	"unfav_Undocumented"
[875]	"unfav_police"	"unfav_socialists"
[877]	"fav_socialists"	"fav_Police"

```
[879] "fav_Police_amongHeardEnough"    "Liberal"
[881] "Conservative"                    "age_group_newbaseline"
[883] "ideo5_newbaseline"               "copartisan"
[885] "quarteryear"                     "econ_quintile"
[887] "culture_quintile"                "immig_quintile"
[889] "racial_quintile"
```

2.3 We subset the data to the first 10 waves

```
# Weekly number of respondents
a %>% group_by(week) %>% tally() %>% head()

# A tibble: 6 x 2
  week      n
  <date>   <int>
1 2019-07-14 3906
2 2019-07-21 5990
3 2019-07-28 6236
4 2019-08-04 6082
5 2019-08-11 5097
6 2019-08-18 7685

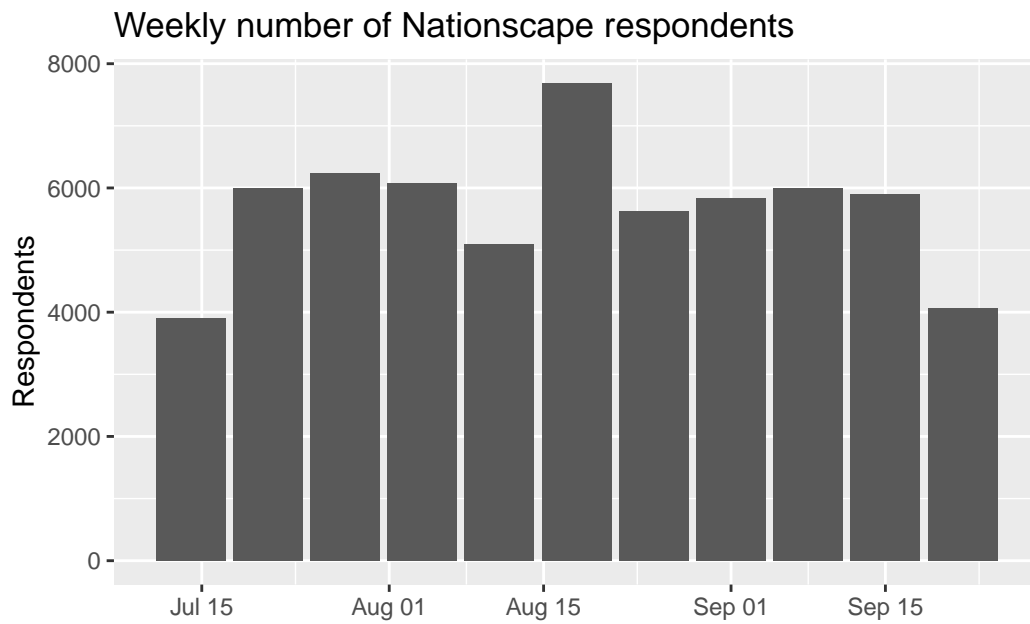
...

a %>% summarise(first = min(date), last = max(date))

# A tibble: 1 x 2
  first      last
  <date>   <date>
1 2019-07-18 2019-09-25
```

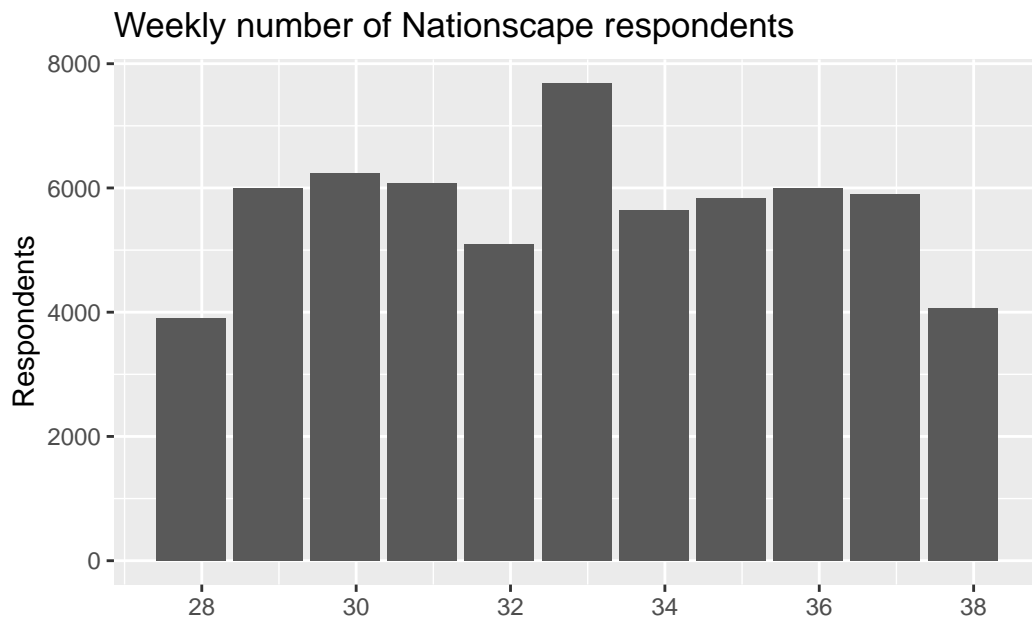
2.4

```
a %>% group_by(week) %>% tally() %>%
  ggplot(aes(x=week,y=n)) + geom_bar(stat="identity") +
  labs(x="",y="Respondents",title="Weekly number of Nationscape respondents")
```



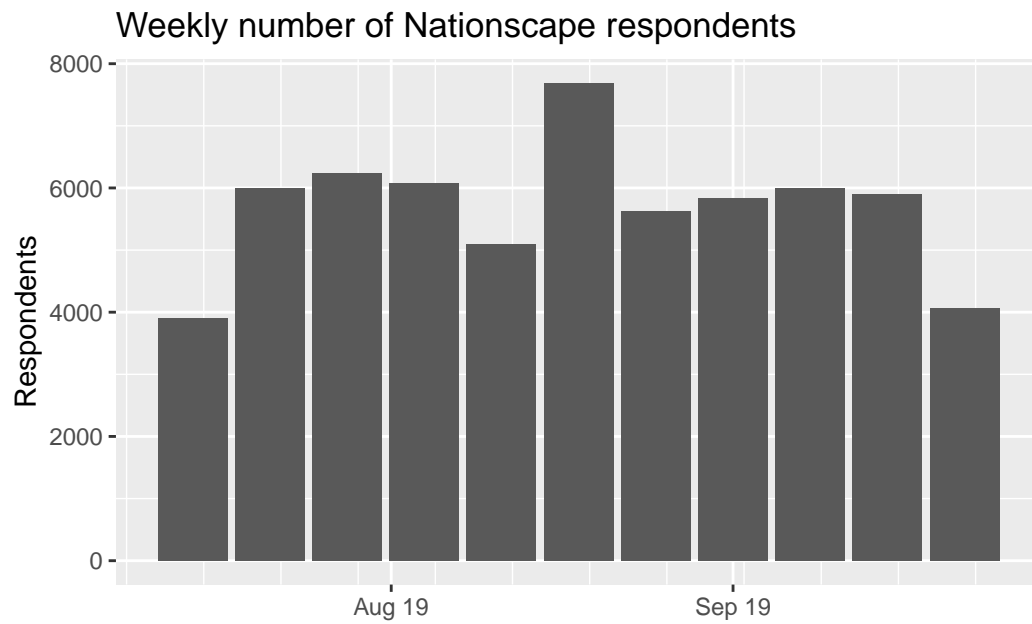
2.5

```
a %>% group_by(week) %>% tally() %>%  
  ggplot(aes(x=week,y=n)) + geom_bar(stat="identity") +  
  labs(x="",y="Respondents",title="Weekly number of Nationscape respondents") +  
  scale_x_date(date_breaks = "2 weeks", date_labels = "%W")
```



2.6

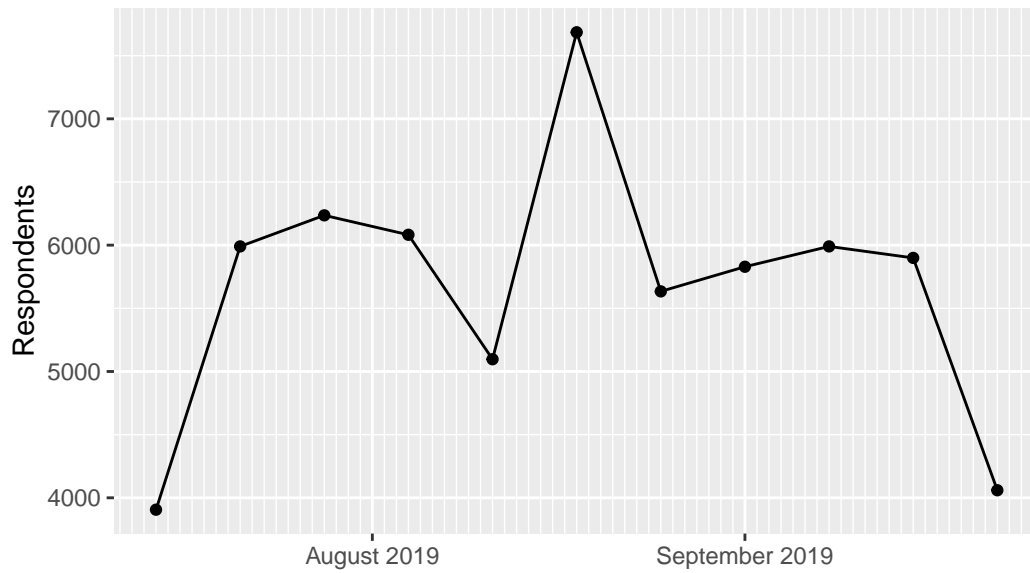
```
a %>% group_by(week) %>% tally() %>%  
  ggplot(aes(x=week,y=n)) + geom_bar(stat="identity") +  
  labs(x="",y="Respondents",title="Weekly number of Nationscape respondents") +  
  scale_x_date(date_breaks = "1 months", minor_breaks = "1 weeks",date_labels = "%b %y")
```



2.7

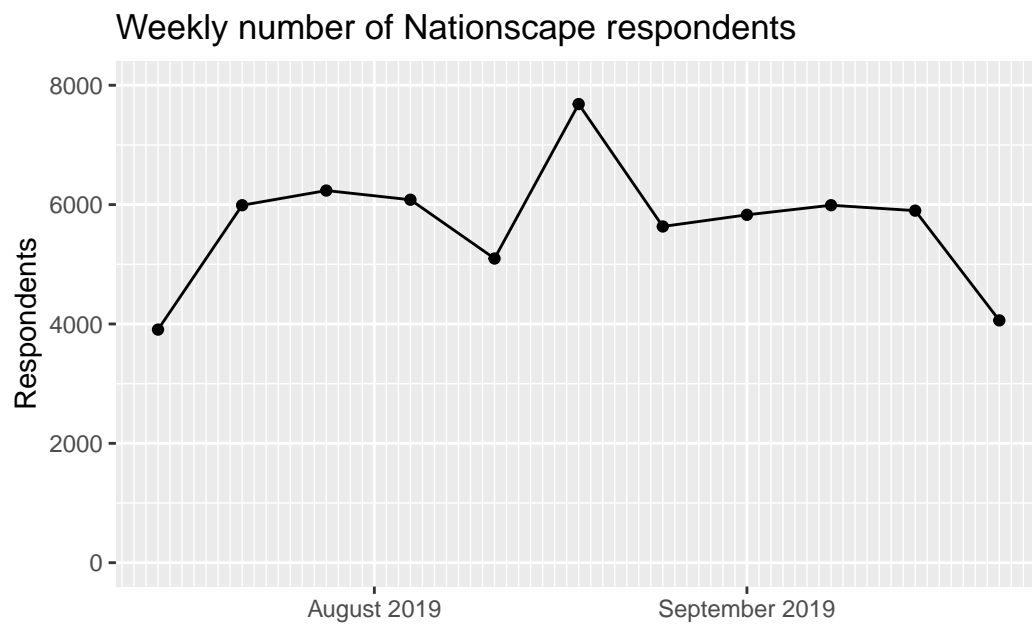
```
a %>% group_by(week) %>% tally() %>%  
  ggplot(aes(x=week,y=n)) + geom_line() + geom_point() +  
  labs(x="",y="Respondents",title="Weekly number of Nationscape respondents") +  
  scale_x_date(date_breaks = "1 months", minor_breaks = "1 days",date_labels = "%B %Y")
```

Weekly number of Nationscape respondents



2.8

```
a %>% group_by(week) %>% tally() %>%  
  ggplot(aes(x=week,y=n)) + geom_line() + geom_point() +  
  labs(x="",y="Respondents",title="Weekly number of Nationscape respondents") +  
  scale_x_date(date_breaks = "1 months", minor_breaks = "1 days",date_labels = "%B %Y") +  
  ylim(c(0,8000))
```



2.9

2.9.1 Favorability of AOC

```
table(a$aoc_Favorable)
```

```

  0      1
19896 16993

```

```
... 
```

```
a %>% summarise(AOC_fav = mean(aoc_Favorable))
```

```

# A tibble: 1 x 1
  AOC_fav
  <dbl>
1      NA

```

```
... 
```



```

a %>% summarise(AOC_fav = mean(aoc_Favorable,na.rm=T),

                 AOC_fav_w = weighted.mean(aoc_Favorable,
                                             w=weight,
                                             na.rm=T))

# A tibble: 1 x 2
  AOC_fav AOC_fav_w
  <dbl>    <dbl>
1  0.461    0.445

```

2.10

Favorability of AOC

```
pollster::topline(a,aoc_Favorable,weight = weight)
```

```

Warning: There was 1 warning in `mutate()`.
i In argument: `aoc_Favorable = forcats::fct_explicit_na(aoc_Favorable)`.
Caused by warning:
! `fct_explicit_na()` was deprecated in forcats 1.0.0.
i Please use `fct_na_value_to_level()` instead.
i The deprecated feature was likely used in the pollster package.
  Please report the issue to the authors.

```

```

# A tibble: 3 x 5
  Response Frequency Percent `Valid Percent` `Cumulative Percent`
  <fct>      <dbl>    <dbl>         <dbl>         <dbl>
1 0          20364.    32.6          55.5          55.5
2 1          16303.    26.1          44.5          100
3 (Missing)   25741.    41.2           NA           NA

```

...

```

pollster::crosstab(df = a,
                  x = gender,
                  y = aoc_Favorable,weight = weight)

```

```
# A tibble: 2 x 4
  gender   `0`   `1`     n
  <fct> <dbl> <dbl> <dbl>
1 Female  51.0  49.0 17352.
2 Male    59.6  40.4 19315.

...

pollster::crosstab(df = a, x = gender,
                   y = aoc_Favorable, weight = weight, format = "long")
```

```
# A tibble: 4 x 4
  gender aoc_Favorable pct     n
  <fct> <fct>          <dbl> <dbl>
1 Female 0           51.0 17352.
2 Female 1           49.0 17352.
3 Male   0           59.6 19315.
4 Male   1           40.4 19315.
```

2.11 Favorability by vote choice (intent)

```
pollster::crosstab(df = a, x = trump_biden,
                   y = aoc_Favorable, weight = weight, format = "long")
```

```
# A tibble: 6 x 4
  trump_biden aoc_Favorable pct     n
  <fct>      <fct>          <dbl> <dbl>
1 Joe Biden  0           28.7 18881.
2 Joe Biden  1           71.3 18881.
3 Donald Trump 0           88.8 14695.
4 Donald Trump 1           11.2 14695.
5 Don't Know 0           61.3  2989.
6 Don't Know 1           38.7  2989.
```

2.12 Favorability by vote choice (intent)

```
pollster::crosstab(df = a, x = trump_biden,  
                  y = aoc_Favorable, weight = weight, format = "long") %>%  
  filter(aoc_Favorable==1)
```

A tibble: 3 x 4

	trump_biden	aoc_Favorable	pct	n
	<fct>	<fct>	<dbl>	<dbl>
1	Joe Biden	1	71.3	18881.
2	Donald Trump	1	11.2	14695.
3	Don't Know	1	38.7	2989.

2.13 Favorability by vote choice (intent)

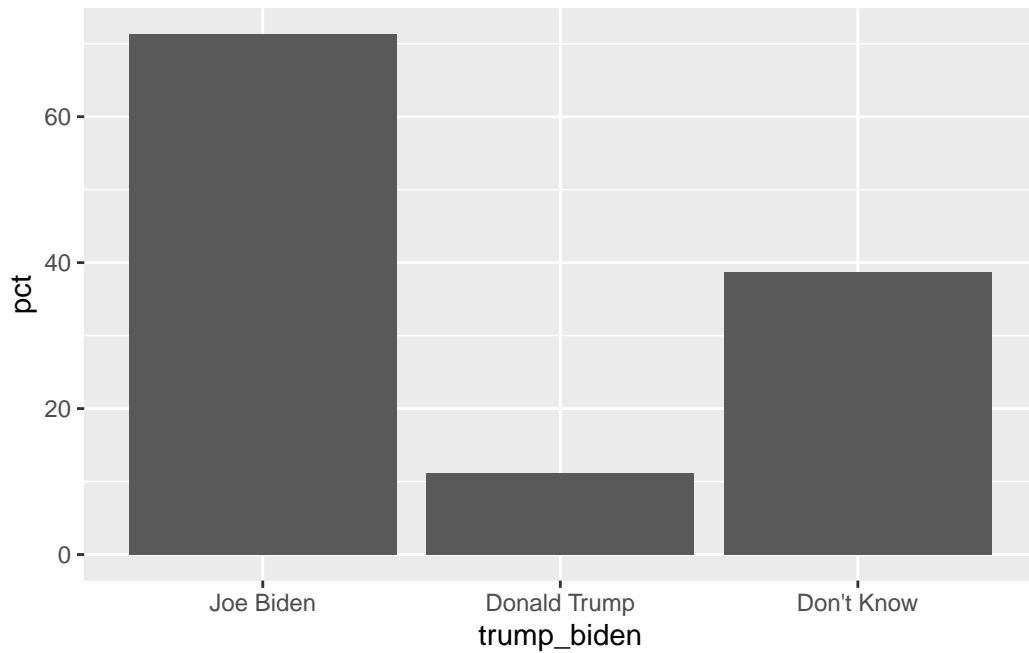
```
pollster::crosstab(df = a, x = trump_biden,  
                  y = aoc_Favorable, weight = weight, format = "long") %>%  
  filter(aoc_Favorable==1)  
# ggplot(aes(x= ..., y = ...))
```

2.14 Favorability by vote choice (intent)

```
pollster::crosstab(df = a, x = trump_biden,  
                  y = aoc_Favorable, weight = weight, format = "long") %>%  
  filter(aoc_Favorable==1) %>%  
  ggplot(aes(x= trump_biden, y = pct)) + geom_col()
```

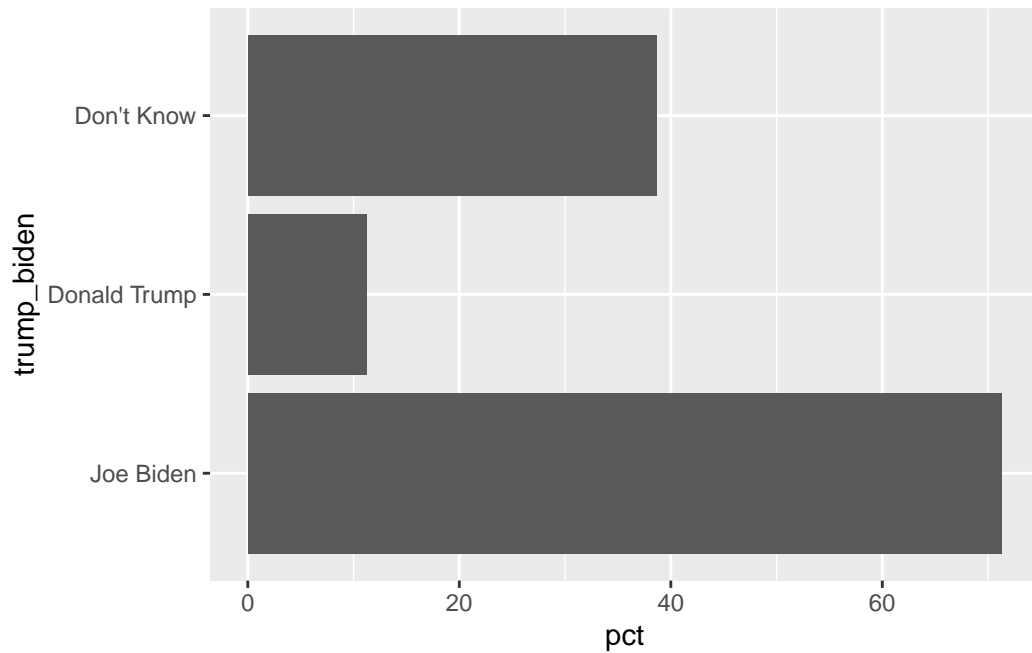
2.15 Favorability by vote choice (intent)

```
pollster::crosstab(df = a, x = trump_biden,  
                  y = aoc_Favorable, weight = weight, format = "long") %>%  
  filter(aoc_Favorable==1) %>%  
  ggplot(aes(x= trump_biden, y = pct)) + geom_col()
```



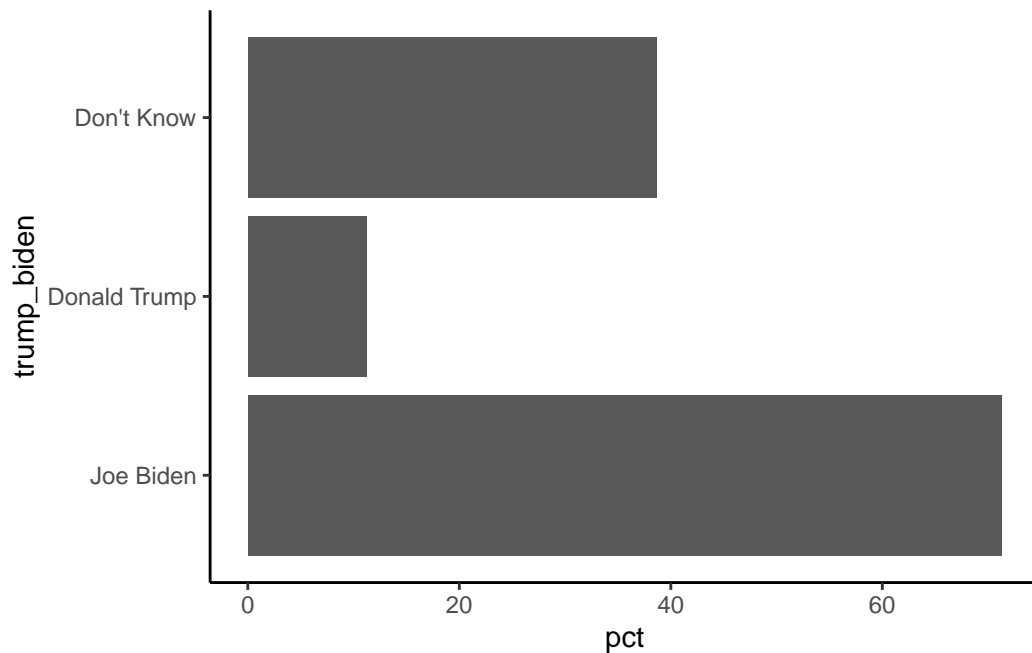
2.16 Favorability by vote choice (intent)

```
pollster::crosstab(df = a, x = trump_biden,  
                   y = aoc_Favorable, weight = weight, format = "long") %>%  
  filter(aoc_Favorable==1) %>%  
  ggplot(aes(y= trump_biden, x = pct)) + geom_col()
```



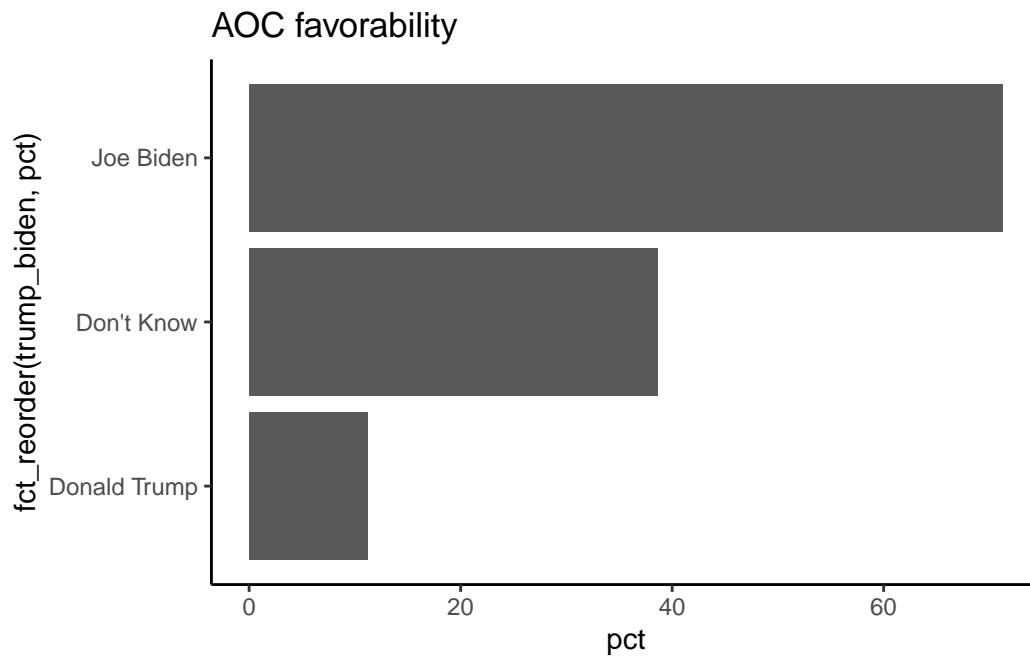
2.17 Favorability by vote choice (intent)

```
pollster::crosstab(df = a, x = trump_biden,  
                   y = aoc_Favorable, weight = weight, format = "long") %>%  
  filter(aoc_Favorable==1) %>%  
  ggplot(aes(y= trump_biden, x = pct)) + geom_col() + theme_classic()
```



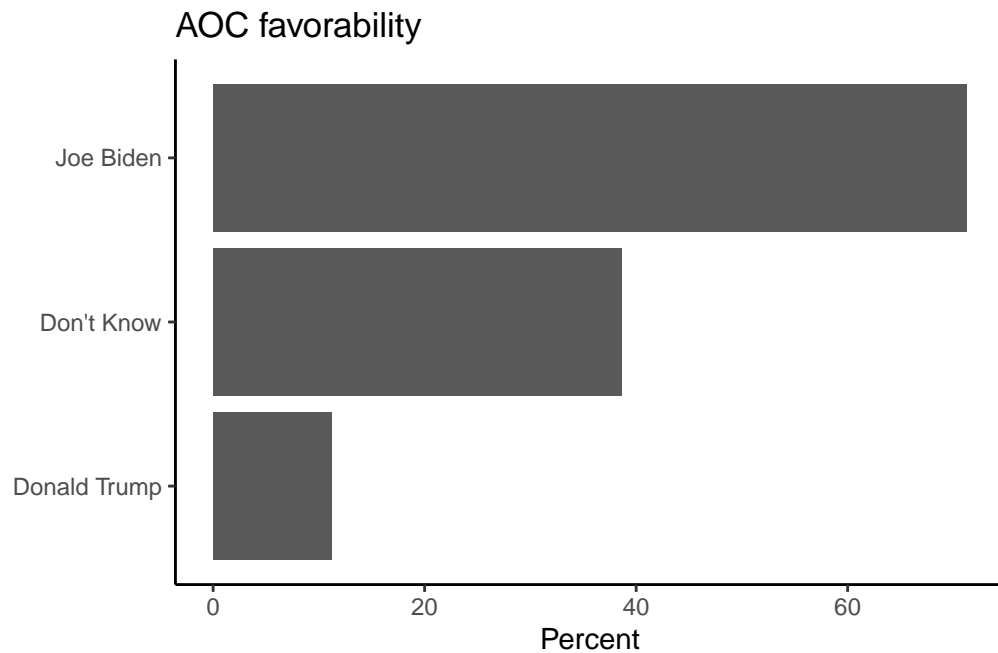
2.18 Favorability by vote choice (intent)

```
pollster::crosstab(df = a, x = trump_biden,  
                   y = aoc_Favorable, weight = weight, format = "long") %>%  
  filter(aoc_Favorable==1) %>%  
  ggplot(aes(y= fct_reorder(trump_biden,pct),  
             x = pct)) + geom_col() + theme_classic() + ggtitle("AOC favorability")
```



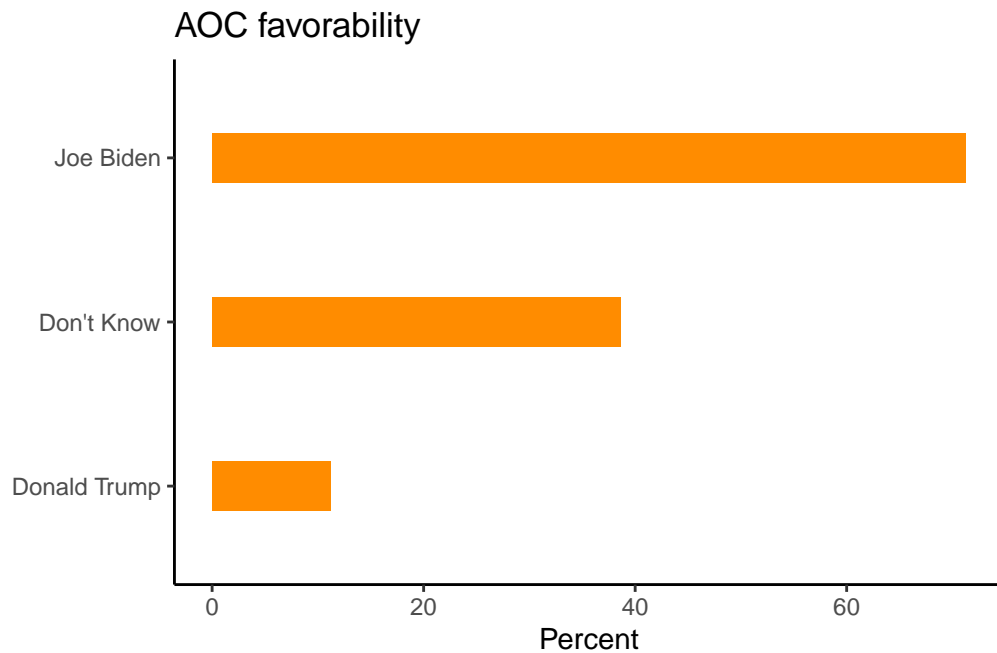
2.19 Favorability by vote choice (intent)

```
pollster::crosstab(df = a, x = trump_biden,
                   y = aoc_Favorable, weight = weight, format = "long") %>%
  filter(aoc_Favorable==1) %>%
  ggplot(aes(y= fct_reorder(trump_biden,pct),
               x = pct)) + geom_col() + theme_classic() +
  labs(x="Percent", y = "", title = "AOC favorability")
```



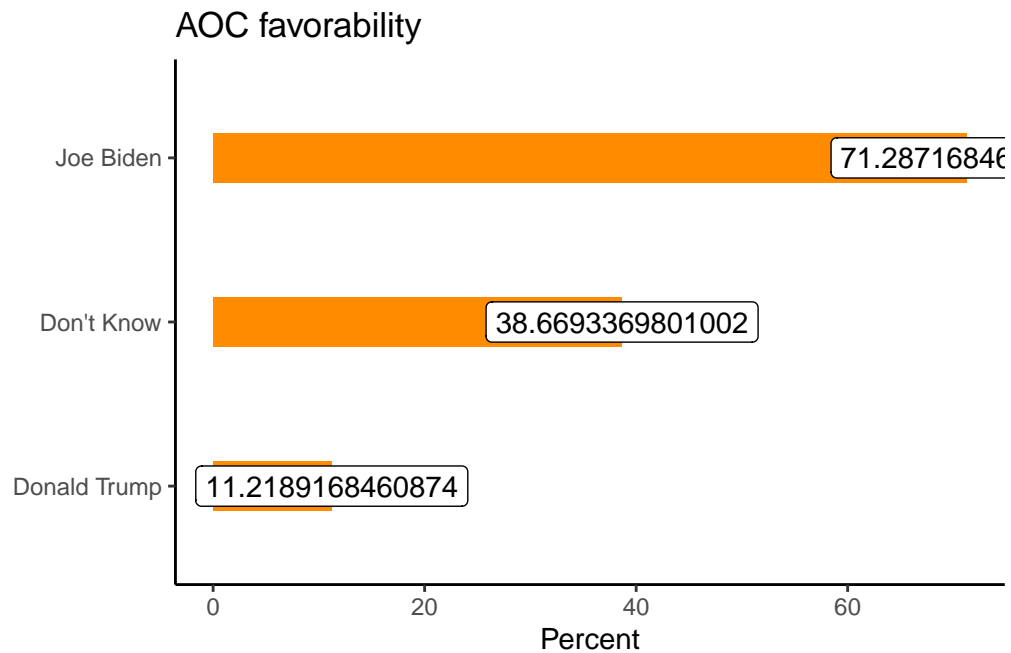
2.20 Favorability by vote choice (intent)

```
pollster::crosstab(df = a, x = trump_biden,  
                   y = aoc_Favorable, weight = weight, format = "long") %>%  
  filter(aoc_Favorable==1) %>%  
  ggplot(aes(y= fct_reorder(trump_biden,pct),  
             x = pct)) + geom_col(width = .3, fill="darkorange") + theme_classic() +  
  labs(x="Percent", y = "", title = "AOC favorability")
```

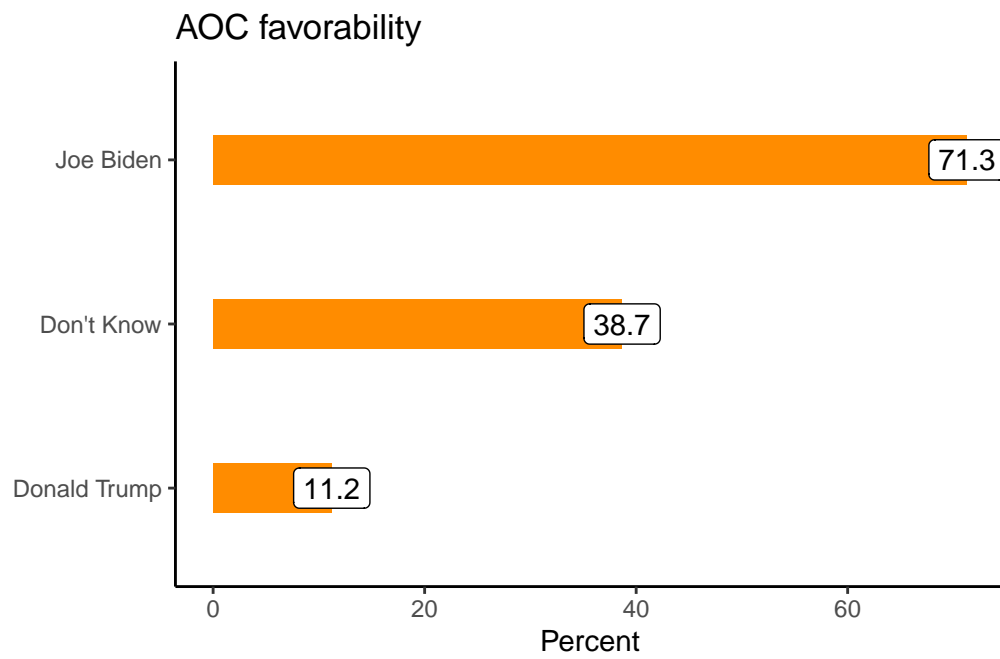
2.21 Favorability by vote choice (intent)

```
pollster::crosstab(df = a, x = trump_biden,  
                   y = aoc_Favorable, weight = weight, format = "long") %>%  
  filter(aoc_Favorable==1) %>%  
  ggplot(aes(y= fct_reorder(trump_biden,pct),  
             x = pct)) + geom_col(width = .3, fill="darkorange") + theme_classic() +  
  labs(x="Percent", y = "", title = "AOC favorability") +  
  geom_label(aes(label=pct))
```



2.22 Favorability by vote choice (intent)

```
pollster::crosstab(df = a, x = trump_biden,
                  y = aoc_Favorable, weight = weight, format = "long") %>%
  filter(aoc_Favorable==1) %>%
  ggplot(aes(y= fct_reorder(trump_biden,pct),
                x = pct)) + geom_col(width = .3, fill="darkorange") + theme_classic() +
  labs(x="Percent", y = "", title = "AOC favorability") +
  geom_label(aes(label=round(pct,1)))
```



3 AJPS 2021 dataset

```
d1
```

```
# A tibble: 2,000 x 94
  caseid female edu black hispanic age income pid ideo interest attend
  <chr>    <dbl> <dbl> <dbl>    <dbl> <dbl> <dbl> <dbl> <dbl>    <dbl> <dbl>
1 R_24COU~      1     5     0        0    23     2     7     3        5     3
2 R_2B2nP~      0     6     0        0    39     7     4     6        3     2
3 R_p5eQb~      0     3     0        0    43     4     4     1        3     2
4 R_2dYYB~      0     2     1        0    22     2     1     7        4     3
5 R_3sgIL~      0     3     1        0    40     5     1     1        4     3
6 R_31Ab1~      0     6     0        0    28     4     4     4        3     1
7 R_2f36X~      0     6     0        0    41     7     4     2        4     2
8 R_2XcYI~      0     2     1        0    21     4     1     3        4     4
9 R_339E8~      1     6     0        0    58     6     3     3        4     4
10 R_3mlfI~      0     5     0        0    43     6     1     1        5     4
# i 1,990 more rows
# i 83 more variables: facebook <dbl+lbl>, twitter <dbl+lbl>, reddit <dbl+lbl>,
#   chans <dbl>, con1 <dbl>, con2 <dbl>, con3 <dbl>, con4 <dbl>, conwis <dbl>,
#   msm <dbl>, onepercent <dbl>, deepstate <dbl>, goodevil <dbl+lbl>,
#   vio1 <dbl>, vio2 <dbl>, violence <dbl>, argue1 <dbl>, argue2 <dbl>,
#   argue3 <dbl>, argument <dbl>, pop1 <dbl>, pop2 <dbl>, official <dbl>,
#   manip1 <dbl>, manip2 <dbl>, manip3 <dbl>, manip4 <dbl>, ...
```

```
table(d2$climatechange)
```

```
 1    2    3    4    5
733 454 395 233 206
```

```
table(d2$climatechangeBIN)
```

	0	1
	1582	439

```
d2 %>% count(climatechangeBIN)
```

```
# A tibble: 3 x 2
  climatechangeBIN     n
      <dbl> <int>
1             0 1582
2             1  439
3            NA     2
```

Are the missing observations the same for the original and the recoded variable? (If not, we would want to check whether earlier code did something unintended.)

```
d2 %>% count(climatechangeBIN, climatechange)
```

```
# A tibble: 6 x 3
  climatechangeBIN climatechange     n
      <dbl>          <dbl> <int>
1             0             1   733
2             0             2   454
3             0             3   395
4             1             4   233
5             1             5   206
6            NA            NA     2
```

4 Economic data over time

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr      1.1.1      v readr      2.1.4
v forcats    1.0.0      v stringr    1.5.0
v ggplot2    3.4.2      v tibble     3.2.1
v lubridate  1.9.2      v tidyr      1.3.0
v purrr      1.0.1

-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()     masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
```

```
D <- read_csv("data_macro/inflation_WDI.csv")
```

```
Rows: 8522 Columns: 4
-- Column specification -----
Delimiter: ","
chr (2): country, iso3c
dbl (2): inflation, year

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
oecd <- read_csv("data_macro/oecd_codes.csv")
```

```
Rows: 35 Columns: 2
-- Column specification -----
Delimiter: ","
chr (1): iso3c
```

```
dbl (1): OECD
```

- i Use ``spec()`` to retrieve the full column specification for this data.
- i Specify the column types or set ``show_col_types = FALSE`` to quiet this message.

```
#biggerText <- jzPack::biggerText

# Merge in OECD indicators
D <- left_join(D, oecd)
```

Joining with ``by = join_by(iso3c)``

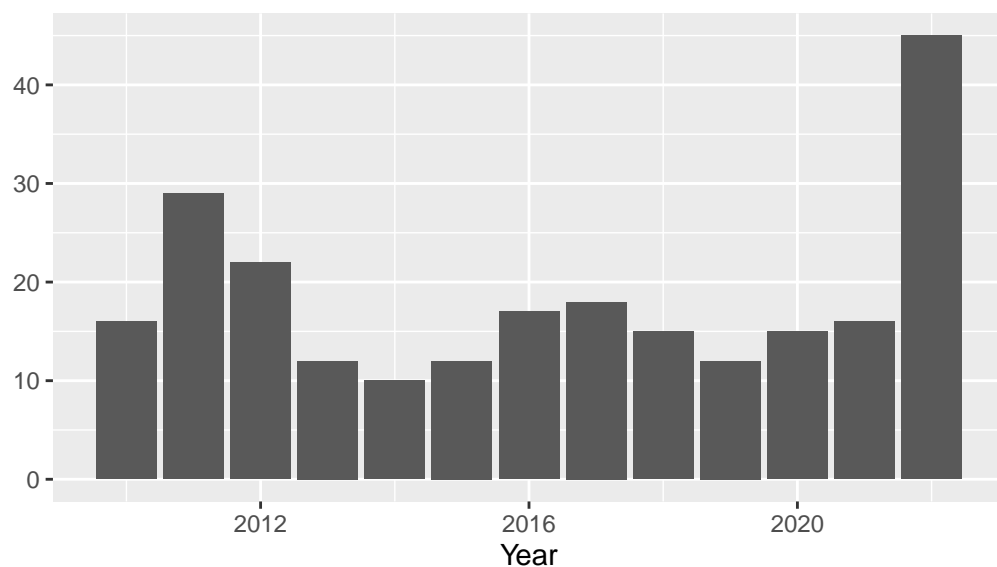
In how many countries has inflation exceeded 10% in 2022:

```
D %>%
  filter(year == 2022) %>%
  mutate(over10 = ifelse(inflation > 10, 1, 0)) %>%
  count(over10)
```

```
# A tibble: 2 x 2
  over10     n
  <dbl> <int>
1      0     82
2      1     45
```

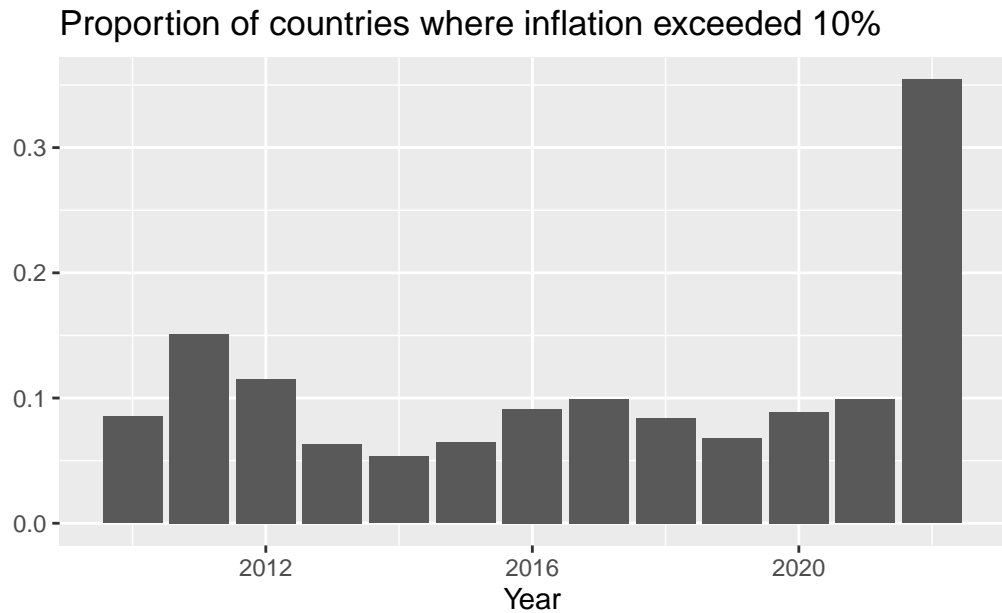
```
D %>%
  filter(year %in% c(2010:2022)) %>%
  mutate(over10 = ifelse(inflation > 10, 1, 0)) %>%
  group_by(year) %>%
  summarise(n = n(),
            over10 = sum(over10)) %>%
  ggplot(aes(x=year, y=over10)) +
  geom_col() +
  labs(y="", x="Year",
       title = "Number of countries with inflation 10%+",
       caption = "Data: WDI")
```

Number of countries with inflation 10%+



Data: WDI

```
D %>%
  filter(year %in% c(2010:2022)) %>%
  mutate(over10 = ifelse(inflation > 10, 1, 0)) %>%
  group_by(year) %>%
  summarise(n = n(),
            prop = sum(over10)/n) %>%
  ggplot(aes(x=year,
            y=prop)) +
  geom_col() +
  labs(y="", x="Year",
       title = "Proportion of countries where inflation exceeded 10%",
       caption = "Data: WDI")
```

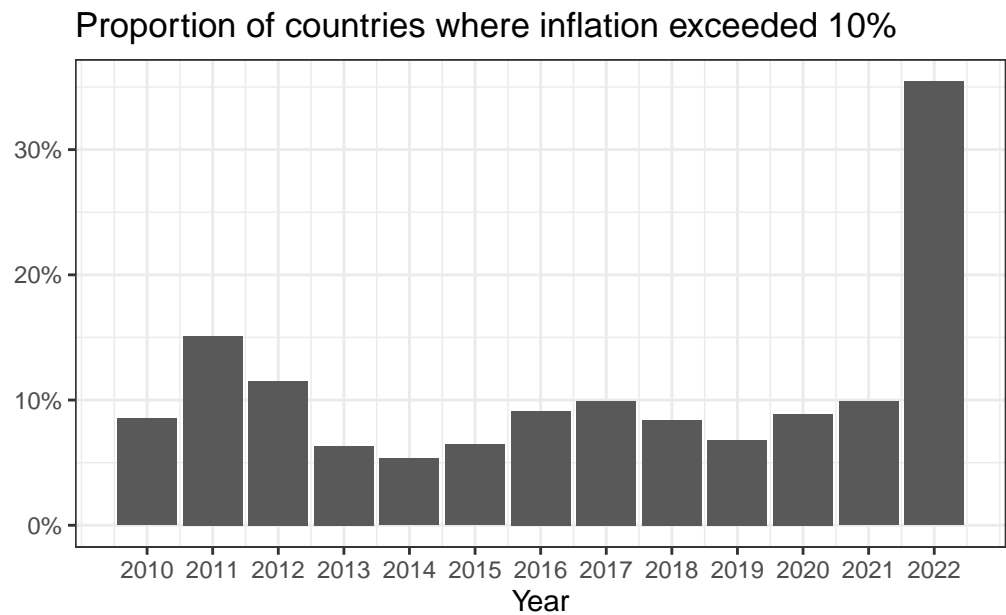



Data: WDI

There are some obvious problems with the x-axis here, so let's solve that, plus a few minor issues:

```
barplot_V2 <- D %>%
  filter(year %in% c(2010:2022)) %>%
  mutate(over10 = ifelse(inflation > 10, 1, 0)) %>%
  group_by(year) %>%
  summarise(n = n(),
            prop = sum(over10)/n) %>%
  ggplot(aes(x=year,
            y=prop)) +
  geom_col() +
  labs(y="", x="Year",
       title = "Proportion of countries where inflation exceeded 10%",
       caption = "Data: WDI") +
  scale_x_continuous(breaks = c(2010:2022)) +
  scale_y_continuous(labels = scales::percent) +
  theme(text = element_text(size=15)) +
  theme_bw()
```

barplot_V2



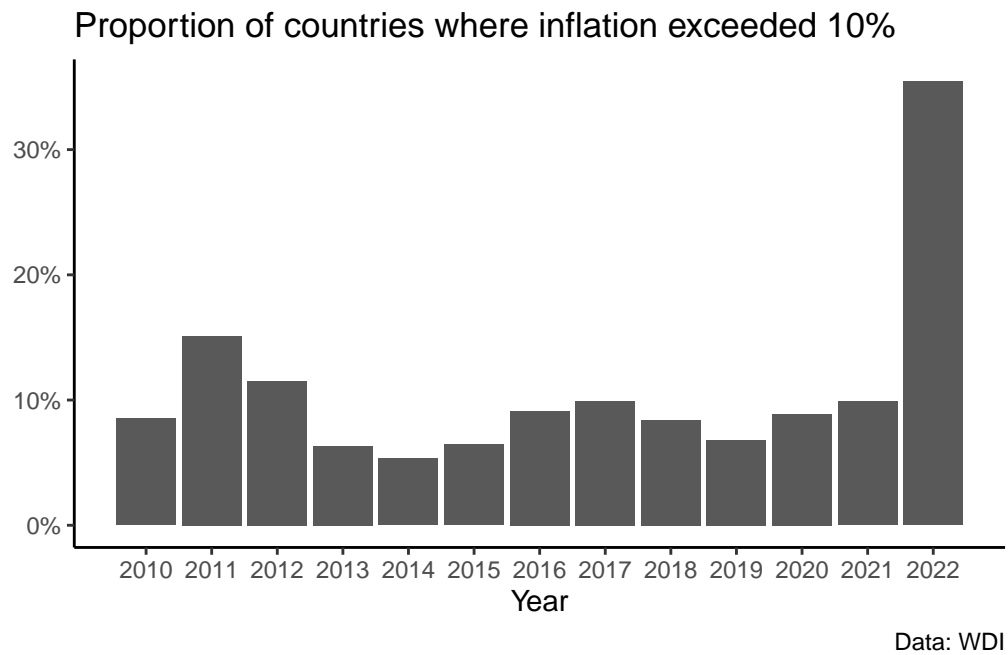
Data: WDI

We improves the labels of both axies, and increased the font size.

The new version is reasonably good, but there are still unnecessary gridlines, right?

You could either use `theme_classic()` instead of `theme_bw()`.

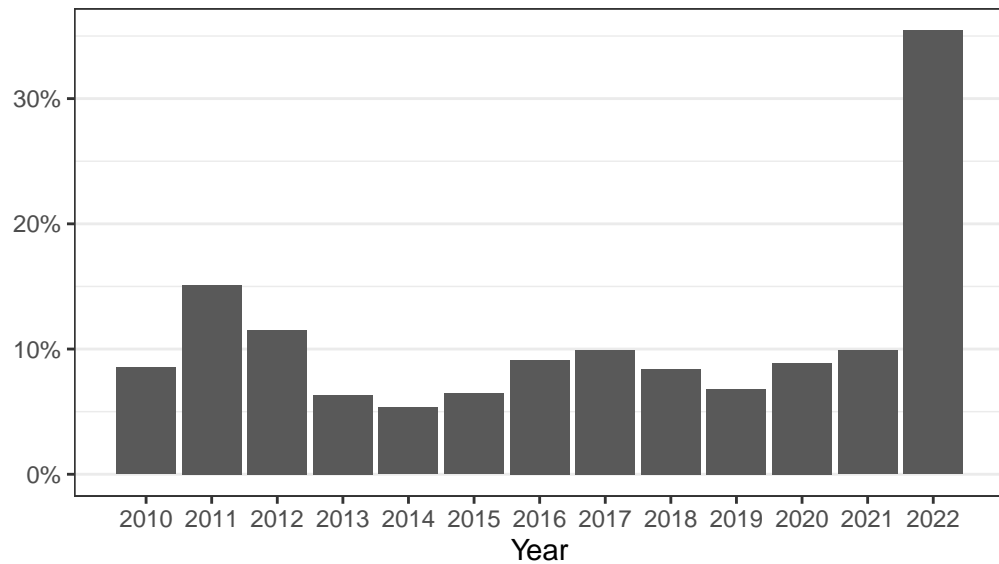
```
barplot_V2 +  
  theme_classic()
```



Or we can keep `theme_bw()` and make some adjustments:

```
barplot_V2 +  
  theme(panel.grid.minor.x = element_blank(),  
        panel.grid.major.x = element_blank())
```

Proportion of countries where inflation exceeded 10%

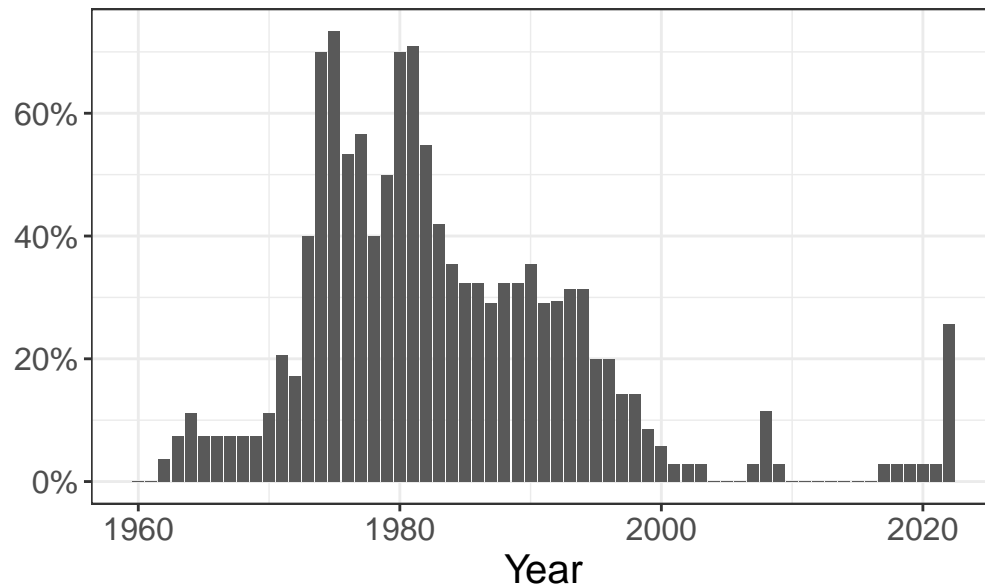


Data: WDI

4.0.1 Zomm in on OECD countries and increase the time horizon

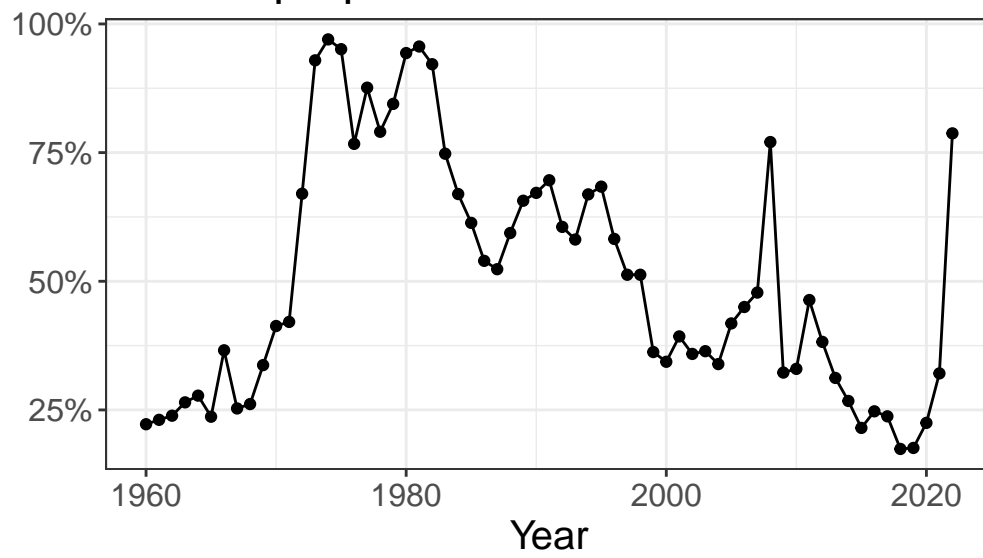
```
D %>% filter(OECD==1) %>%
  filter(!is.na(inflation)) %>%
  mutate(over10 = ifelse(inflation > 10, 1, 0)) %>%
  group_by(year) %>%
  summarise(n = n(),
            prop = sum(over10)/n) %>%
  ggplot(aes(x=year,
            y=prop)) +
  geom_col() +
  theme_bw() +
  theme(text = element_text(size=15)) +
  scale_y_continuous(labels = scales::percent) +
  labs(y="", x="Year",
       title = "Proportion of countries where inflation exceeded 10% (OECD countries only)")
```

Proportion of countries where inflation exce



```
D %>%
  mutate(over5 = ifelse(inflation > 5, 1, 0)) %>%
  group_by(year) %>%
  summarise(n = n(),
            prop = sum(over5)/n) %>%
  ggplot(aes(x=year,
            y=prop)) +
  geom_line() + geom_point() +
  theme_bw() +
  theme(text = element_text(size=15)) +
  scale_y_continuous(labels = scales::percent) +
  labs(y="", x="Year",
       title = "Global proportion of countries where inflation exceeded 5%",
       caption = "Data: WDI.")
```

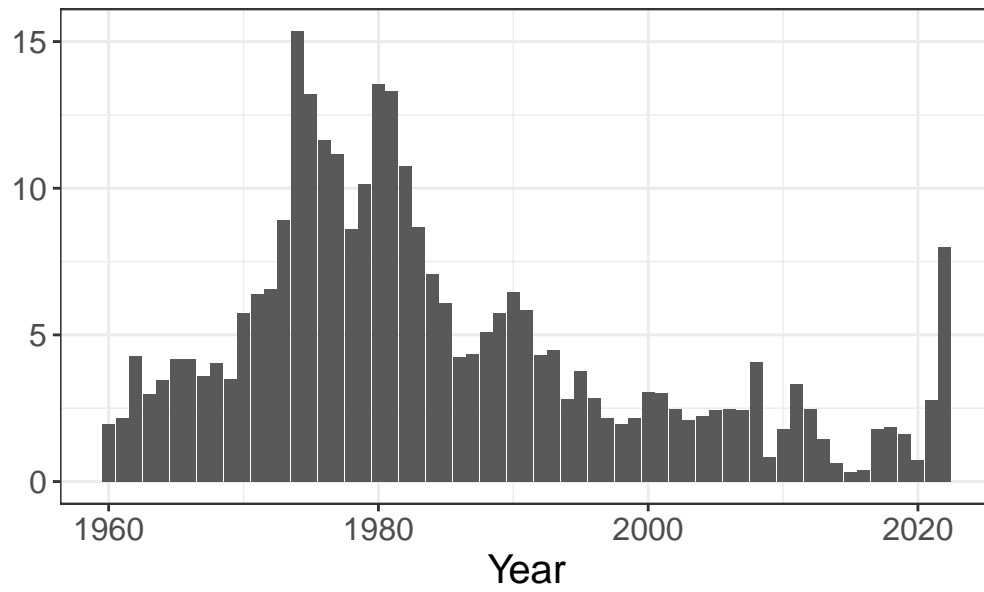
Global proportion of countries where inflat



Data: WDI.

```
D %>% filter(OECD==1) %>%
  filter(!is.na(inflation)) %>%
  group_by(year) %>%
  summarise(M = median(inflation)) %>%
  ggplot(aes(x=year,
             y=M)) +
  # geom_point() + geom_line() +
  geom_col() +
  theme_bw() +
  theme(text = element_text(size=15)) +
  labs(y="", x="Year",
       title = "Median inflation in OECD countries")
```

Median inflation in OECD countries



5 Standard charts

In summary, this book has no content whatsoever.

`1 + 1`

[1] 2

6 Advanced ggplot

[See the lecture notes on the course website]

7 Visualizing statistical models

A more accurate title, of course, would be “visualizing *outputs* from statistical models”.

[COMING SOON]

References

Useful resources include:

[Gestalt Principles](#)

[Gestalt Principles \(Part 2\)](#)

<https://socviz.co/>

<https://ggplot2-book.org/index.html>

<https://cssbook.net/content/chapter06.html>

<https://storymaps.arcgis.com/stories/1e7f582d478a4b99bd0c70fffeac4c8b>

<https://cup.columbia.edu/book/better-data-visualizations/9780231193115>

<https://journals.sagepub.com/doi/pdf/10.1177/15291006211057899>