

R Notebook

Load packages and data

```
suppressMessages(library(dplyr))
suppressMessages(library(kableExtra))

packageVersion('dplyr')
```

```
## [1] '1.1.4'
```

```
packageVersion('kableExtra')
```

```
## [1] '1.4.0'
```

```
R.version
```

```
##
## platform      _
## arch          aarch64-apple-darwin20
## arch          aarch64
## os            darwin20
## system        aarch64, darwin20
## status
## major         4
## minor         4.2
## year          2024
## month         10
## day           31
## svn rev       87279
## language      R
## version.string R version 4.4.2 (2024-10-31)
## nickname      Pile of Leaves
```

```
setwd("~/Desktop/MarshallProjectDataAssignment")
survey_responses <- read.csv("responses.csv")
```

Data exploration

Get acquainted with the data.

```
unique(survey_responses$answer_set) # see what the different answer sets are
```

```
## [1] "all-surveys" "both-surveys"
```

```
nrow(filter(survey_responses, answer_set == "all-surveys"))
```

```
## [1] 21320
```

```
nrow(filter(survey_responses, answer_set == "both-surveys"))
```

```
## [1] 5614
```

The notebook says that “all-surveys” includes the answers across both surveys (set 2024-001 and set 2024-002). The “both-surveys” set may only include individuals who responded to both surveys. I am choosing to analyze the “all-surveys” set for the cursory analysis since it includes more responses and since that is the set that is used by the Marshall Project notebook.

```
df <- filter(survey_responses, answer_set == "all-surveys") # data frame that only includes the "all-answers" survey set.
```

```
df_state <- df[which(df$state != "0_ALL_STATES"),] # get state level data  
df_all_states <- df[which(df$state == "0_ALL_STATES"),] # get data for all states
```

```
print(paste("Number of respondents:", unique(filter(df_all_states, question_ids == "facility_name")$n_respondents)))
```

```
## [1] "Number of respondents: 15481"
```

```
print(paste("Count:", sum(filter(df_all_states, question_ids == "facility_name")$count)))
```

```
## [1] "Count: 15062"
```

15,481 people responded to the facility name question, so it looks like at least 15,481 people in jail responded to the survey nationally.

```
colnames(df) # look at each variable in the dataset
```

```
## [1] "state"
## [2] "answer_set"
## [3] "facility_name"
## [4] "age"
## [5] "count"
## [6] "n_respondents"
## [7] "pct_of_respondents"
## [8] "question_ids"
## [9] "biden_presidency_opinion"
## [10] "climate_change_effects"
## [11] "education"
## [12] "eligible_to_vote"
## [13] "facility_type"
## [14] "following_trump_stormy_daniels_trial"
## [15] "gender"
## [16] "independent_preference_for_dem_rep"
## [17] "kamala_opinion"
## [18] "kamala_record_familiarity"
## [19] "kamala_record_on_crime"
## [20] "labor_unions_opinion"
## [21] "likely_to_vote_in_november_election"
## [22] "party"
## [23] "people_with_convictions_should_vote"
## [24] "race"
## [25] "sentence_length"
## [26] "sentence_so_far_served"
## [27] "skills_for_politics"
## [28] "thought_to_november_election"
## [29] "trump_dictator_first_day_opinion"
## [30] "trump_punishment"
## [31] "us_ready_elect_woman_president"
## [32] "vote_impact_government_level"
## [33] "who_vote_for"
## [34] "who_vote_for_supplement"
```

The survey question “people_with_convictions_should_vote” appears to be the most closely related to our question about incarcerated people’s opinions on increasing voting rights for the current and formerly incarcerated, so I will concentrate on responses to this question first.

People with convictions should vote survey responses

```
responses <- filter(df_state, question_ids == "facility_name__people_with_convictions_sh
ould_vote") # data set with responses to 'people_with_convictions_should_vote'
unique(responses$people_with_convictions_should_vote) # get the different responses we s
ee to this question
```

```
## [1] "While incarcerated for any crime"
## [2] "After they have left prison"
## [3] "While incarcerated, but only if they were convicted of non-violent offenses"
```

Here are the different responses we see to the “people_with_convictions_should_vote” question.

```
facility_count <- data.frame(aggregate(responses$facility_name, list(responses$n_respondents, responses$state), FUN = min)) # get number of respondents by facility
colnames(facility_count) <- c("n_respondents", "state", "facility_name") # change column names
sum(facility_count$n_respondents)
```

```
## [1] 375
```

There were 375 respondents to this question across all the facilities surveyed. This is low compared to the number of people who responded to other questions. The limited number of respondents could be problematic for our story.

```
sum(responses$count)
```

```
## [1] 330
```

We can see 330 responses according to the count variable, but it looks like the total number of respondents is 375. The discrepancy is possibly because we cannot see answer groups with less than 5 people. However, it's possible that n_respondents does counts a person who took both surveys as two responses. According to the notebook, it looks like the explanation is the former, but I need to clarify this. I will assume that the count is different from the number of respondents because some answers are suppressed for now.

Here I'm going to generate percentage breakdowns of the unique answers to the voting rights questions based on the total number of responses. Note that this is likely not the true percent of people who chose each of these answers since we may not be able see all of the responses / the true counts.

```
# calculate
unique_responses <- unique(responses$people_with_convictions_should_vote)
for(response in unique_responses){
  total <- sum(filter(responses, people_with_convictions_should_vote == response)$count)
  pct <- (total / 375)*100 # using 375 here because that's the total number of responses
  print(paste(total, "people /", round(pct, digits = 1), "percent of responses said that
people with felony convictions should vote:", response))
}
```

```
## [1] "213 people / 56.8 percent of responses said that people with felony convictions
should vote: While incarcerated for any crime"
## [1] "95 people / 25.3 percent of responses said that people with felony convictions s
hould vote: After they have left prison"
## [1] "22 people / 5.9 percent of responses said that people with felony convictions sh
ould vote: While incarcerated, but only if they were convicted of non-violent offenses"
```

The majority of the respondents think people with felony convictions should be able to vote while incarcerated for any crime.

```
(sum(filter(responses, people_with_convictions_should_vote == "While incarcerated for any crime")$count) + 4) / 375 # percent if 4 more people from 11th facility answered this way
```

```
## [1] 0.5786667
```

We can see the count of people who responded that people with felony convictions should be able to vote while incarcerated for any crime for 10 out of 11 facilities. This means at most 4 more respondents from the 11th facility responded “while incarcerated for any crime” to this question (less than 5 is the threshold for being suppressed). So the true number of respondents who answered in this way ranges from 213 - 217 meaning the true percentage out of all respondents is somewhere between 56.8 - 57.9%

```
print(paste("Number of facilities with responses to the voting rights question:", length(unique(responses$facility_name))))
```

```
## [1] "Number of facilities with responses to the voting rights question: 11"
```

```
print(unique(responses$state))
```

```
## [1] "DC" "TX" "WV"
```

Only people from facilities in D.C., Texas, and West Virginia responded to this question (that we are able to see in the filtered results). This may be problematic for a story with a broad national lens.

```
state_summary <- aggregate(facility_count$n_respondents, list(facility_count$state), FUN = sum) # get number of respondents by state
colnames(state_summary) <- c("state","n_respondents") # change column names for legibility
state_summary
```

state	n_respondents
<chr>	<int>
DC	12
TX	105
WV	258
3 rows	

About 69% of people who responded to this question were incarcerated in jails in West Virginia. This means our sample is heavily skewed; this doesn't give us a sense of the opinions of people in jails elsewhere in the country.

```
kable(df_state %>% select(state, facility_name, race, count, n_respondents, pct_of_respondents, question_ids) %>% filter(state == "WV", question_ids == "facility_name__race", race == "White") %>% arrange(facility_name))
```

state	facility_name	race	count	n_respondents	pct_of_respondents	question_ids
WV	WV DCR Central Regional Jail	White	29	39	0.7435897	facility_name__race
WV	WV DCR Eastern Regional Jail	White	12	20	0.6000000	facility_name__race
WV	WV DCR North Central Regional Jail	White	77	115	0.6695652	facility_name__race
WV	WV DCR Potomac Highlands Regional Jail	White	13	17	0.7647059	facility_name__race
WV	WV DCR South Central Regional Jail	White	27	35	0.7714286	facility_name__race
WV	WV DCR Southern Regional Jail	White	19	27	0.7037037	facility_name__race
WV	WV DCR Southwestern Regional Jail	White	36	47	0.7659574	facility_name__race
WV	WV DCR Tygart Valley Regional Jail	White	49	59	0.8305085	facility_name__race
WV	WV DCR Western Regional Jail	White	28	41	0.6829268	facility_name__race

The majority of respondents from West Virginia facilities who indicated what race they identify as said they were White. But we don't know the racial breakdown of those who answered the voting rights question. However, the evidence we have tells us that responses from West Virginia jails may not capture many opinions from people of other races. This is a potential limitation that we need to look into more.

```
# rearrange factor levels - this is just to control how it is organized in the dataframe
responses$people_with_convictions_should_vote <- factor(responses$people_with_conviction
s_should_vote,
                                                    levels = c("While incarcerated f
or any crime","After they have left prison",
                                                    "While incarcerated,
but only if they were convicted of non-violent offenses"))
# visualize answers from respondents aggregated by facility
kable(responses %>% select(facility_name, people_with_convictions_should_vote, pct_of_re
spondents, count) %>% arrange(people_with_convictions_should_vote, -pct_of_respondents))
```

facility_name	people_with_convictions_should_vote	pct_of_respondents	count
DC DOC: Central Detention Facility	While incarcerated for any crime	0.9166667	11
MTC: Giles Dalby, TX	While incarcerated for any crime	0.6404494	57
WV DCR North Central Regional Jail	While incarcerated for any crime	0.6351351	47

facility_name	people_with_convictions_should_vote	pct_of_respondents	count
WV DCR Potomac Highlands Regional Jail	While incarcerated for any crime	0.6250000	10
WV DCR Southwestern Regional Jail	While incarcerated for any crime	0.6111111	22
WV DCR Western Regional Jail	While incarcerated for any crime	0.6071429	17
WV DCR South Central Regional Jail	While incarcerated for any crime	0.5833333	14
WV DCR Tygart Valley Regional Jail	While incarcerated for any crime	0.4750000	19
WV DCR Central Regional Jail	While incarcerated for any crime	0.4090909	9
WV DCR Southern Regional Jail	While incarcerated for any crime	0.3888889	7
WV DCR Southern Regional Jail	After they have left prison	0.5555556	10
GEO Group-Kinney County, TX	After they have left prison	0.4375000	7
WV DCR Potomac Highlands Regional Jail	After they have left prison	0.3750000	6
WV DCR Central Regional Jail	After they have left prison	0.2727273	6
WV DCR North Central Regional Jail	After they have left prison	0.2702703	20
MTC: Giles Dalby, TX	After they have left prison	0.2584270	23
WV DCR Tygart Valley Regional Jail	After they have left prison	0.2500000	10
WV DCR Western Regional Jail	After they have left prison	0.2142857	6
WV DCR Southwestern Regional Jail	After they have left prison	0.1944444	7
WV DCR Tygart Valley Regional Jail	While incarcerated, but only if they were convicted of non-violent offenses	0.2000000	8
MTC: Giles Dalby, TX	While incarcerated, but only if they were convicted of non-violent offenses	0.0898876	8
WV DCR North Central Regional Jail	While incarcerated, but only if they were convicted of non-violent offenses	0.0810811	6

For 7 of the 11 facilities with respondents to this question, a majority of respondents answered that people convicted of felonies should be able to vote while incarcerated for any crime. The percent of respondents who indicated that people convicted of felonies should be able to vote while incarcerated for any crime varies widely from about 39% to 92%. 1 facility (WV DCR Southern Regional Jail) had a majority of respondents answer that people convicted of felonies should only be able to vote after they have left prison.

The DC DOC: Central Detention Facility had a notably higher percentage of respondents indicate support for people voting while incarcerated for any crime at 92%. From the notebook, it looks like a majority of the respondents from this facility who answered the question on who they are going to vote for said they'd vote for Donald Trump and many indicated they were independents (37%) or democrats (27%). However, these could have been different respondents from those who answered the question on whether people convicted of felonies should be allowed to vote. Further analysis is needed to draw conclusions about how political affiliations and opinions on voting rights intersect.

Takeaways

- About 57% of people who answered the “people_with_convictions_should_vote” question said that people with felony convictions should be able to vote while incarcerated for any crime. The degree of support among respondents varied by jail, but in 7 out of 11 facilities with respondents a majority of respondents supported voting rights for people who were convicted for a felony while still incarcerated.
- Only 375 people from 11 jails in 3 states / territories answered the “people_with_convictions_should_vote” question, this is a small sample that does not allow us to look at opinions in jails in many states. It also appears that the sample has the potential to heavily overrepresent White respondents.

Next steps

- What are the races and political affiliations of the respondents who answered the question on when people with felony convictions should be allowed to vote? Are there any patterns we can identify?
- Look into why so many of the respondents to the voting rights question were from West Virginia. Did facilities in West Virginia do something differently? Double check that the exact same survey was implemented everywhere.
- Look into why so few respondents answered this question compared to some of the other questions.
- Look into answers to other questions to see if they can tell us anything about respondents' opinions on voting rights and increased voting access for incarcerated people.