

# Tidyverse Create - Dplyr

Lu Beyer

## Overview

The dataset I chose is looking at 2022 Republican candidates who denied the 2020 Presidential election results. The Tidyverse package I chose to organize this data is Dplyr, which is my favorite package for cleaning and manipulating data.

```
library(tidyselect)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
data <- read.csv("fivethirtyeight_election_deniers.csv")
```

First we can use the select function within a Tidyverse pipe to easily narrow down the columns we would like to use. I like to do manipulation into a new variable to easily compare against the original, and for easier troubleshooting when issue arise.

We can select in two different ways. First we can select by explicitly stating the columns we want.

```
a <- data %>%
  select(Candidate, Incumbent, State, Office, Stance, Source)

a %>% head(5)
```

```
##      Candidate Incumbent  State      Office      Stance
## 1    Katie Britt      No Alabama    Senator Fully denied
## 2     Jerry Carl      Yes Alabama Representative Fully denied
## 3    Barry Moore      Yes Alabama Representative Fully denied
## 4    Mike Rogers      Yes Alabama Representative Fully denied
## 5 Robert Aderholt      Yes Alabama Representative Fully denied
##                                     Source
## 1                                     NBC News
```

```
## 2 Congressional roll call, Alabama Political Reporter
## 3 Congressional roll call
## 4 Congressional roll call
## 5 Congressional roll call
```

We can also use select to remove columns from our dataframe, and preserve the others.

```
a <- a %>%
  select(-Source)
a %>% head(5)
```

```
##      Candidate Incumbent State Office Stance
## 1    Katie Britt      No Alabama Senator Fully denied
## 2     Jerry Carl     Yes Alabama Representative Fully denied
## 3     Barry Moore     Yes Alabama Representative Fully denied
## 4     Mike Rogers     Yes Alabama Representative Fully denied
## 5 Robert Aderholt     Yes Alabama Representative Fully denied
```

We can use mutate to conditionally modify existing columns. We can also use mutate to conditionally create new columns as well.

Here I will use grepl to search for specific stances to combine into more broad categories.

```
b <- a %>%
  mutate(Stance = ifelse(grepl("accept", Stance, ignore.case = TRUE), "Accept", Stance),
         Stance = ifelse(grepl("denied", Stance, ignore.case = TRUE), "Deny", Stance),
         Stance = ifelse(grepl("avoid|no comment|question", Stance, ignore.case = TRUE), "Neither", Stance))
b %>% head(5)
```

```
##      Candidate Incumbent State Office Stance
## 1    Katie Britt      No Alabama Senator Deny
## 2     Jerry Carl     Yes Alabama Representative Deny
## 3     Barry Moore     Yes Alabama Representative Deny
## 4     Mike Rogers     Yes Alabama Representative Deny
## 5 Robert Aderholt     Yes Alabama Representative Deny
```

## Conclusion

Dplyr is an extremely versatile and useful tool in data cleaning. It easily allows you to transform data into usable data, while maintaining human legibility. Mutate specifically is one of my favorite functions, because it allows you to have a lot of control over how the data looks, and helps you steer the data into how you want to display it for future visualization.