Data Cleaning and Visualization

2024-09-12

```
knitr::opts_chunk$set(echo = TRUE)
library(patchwork)

knitr::opts_chunk$set(
  fig.width = 6,
  fig.asp = .6,
  out.width = "90%"
)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- 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 errors
```

```
library(janitor)
```

```
##
## Attaching package: 'janitor'
##
## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

```
library(readxl)
library(viridisLite)
```

Uploading Milbank Primary Care Data

```
pc_data = readxl::read_excel("data/primary-care-data-2010-2021.xlsx", sheet = "primary-care-data", rang
  janitor::clean_names(
  ) |>
  mutate(across(where(is.character), ~na_if(., "NA"))) |>
  mutate_if(is.character, as.numeric) %>%
  pivot_longer(
    cols = -year,  # All columns except 'year'
    names_to = c("state", "measure"),  # Names of new columns
    names_pattern = "^(..)_?(.*)$"
  )
```

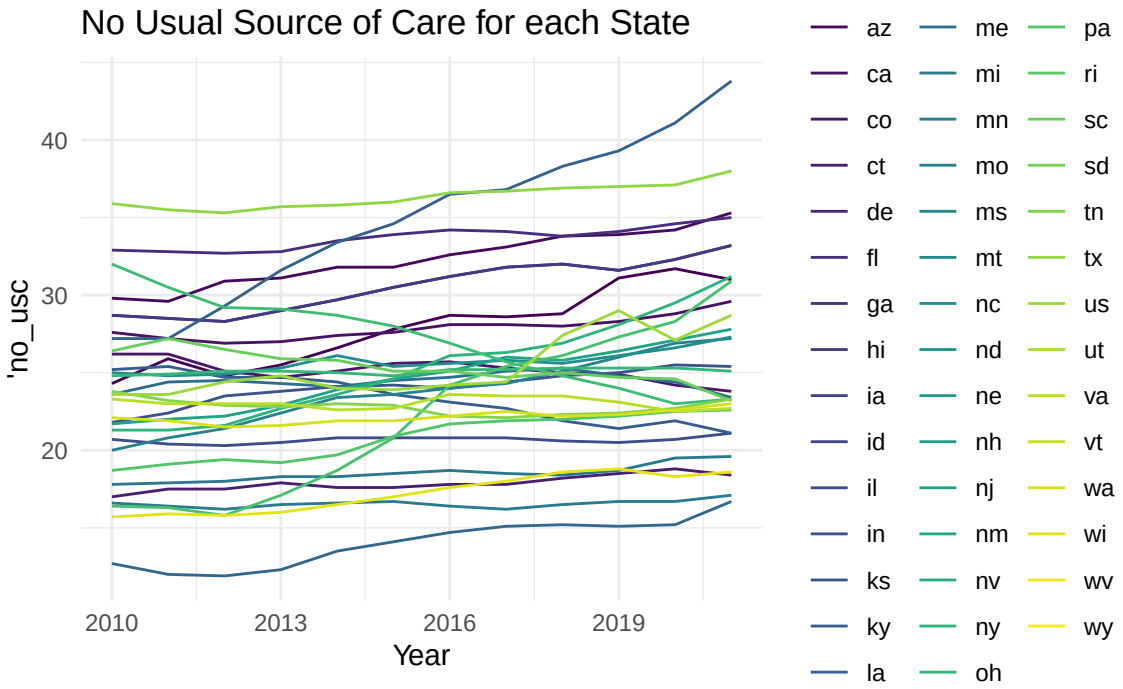
```
## New names:
```

```
## * `AK_no_usc` -> `AK_no_usc...104`
## * `AK_no_usc` -> `AK_no_usc...106`
```

```
no_usc = pc_data |>
  filter(measure == "no_usc")

ggplot(data = no_usc, aes(x = year, y = value, color = state, group = state)) +
  geom_line() +
  scale_color_viridis_d() +  # Use Viridis palette
  labs(title = "No Usual Source of Care for each State",
       x = "Year",
       y = "'no_usc",
       color = "State") +
  theme_minimal()
```

```
## Warning: Removed 216 rows containing missing values or values outside the scale range
## (`geom_line()`).
```



Extracting PCP ratio data

```
pc_ratio = readxl::read_excel("data/primary-care-data-2010-2021.xlsx", sheet = "pcp_ratio_data", range =
  janitor::clean_names() |>
  mutate(across(where(is.character), ~na_if(., "NA"))) |>
  mutate_if(is.character, as.numeric) |>
  pivot_longer(
    cols = -year,  # All columns except 'year'
    names_to = c("state", "measure"),  # Names of new columns
    names_pattern = "^(..)_?(.*)$"
  )

pcp= pc_ratio |>
  filter(measure == "pcp")
```