unit 2

Julia English

2022-09-20

```
library(foreign)
library(tidyverse)
```

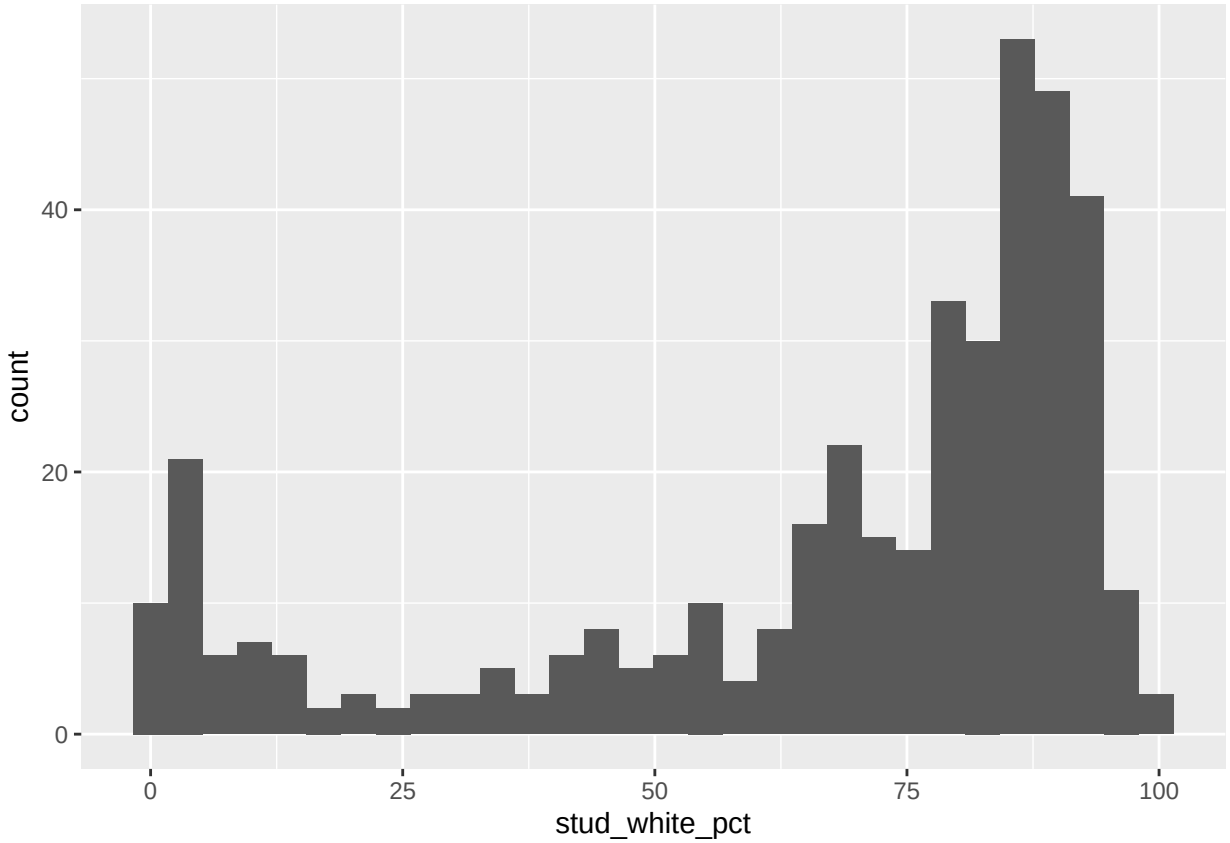
```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.2      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggplot2)
madese <- read.dta('madese.dta')
```

```
w_plot <- madese %>%
  select(stud_white_pct) %>%
  ggplot(mapping = aes(x = stud_white_pct)) +
  geom_histogram()

w_plot
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 4 rows containing non-finite values (stat_bin).
```

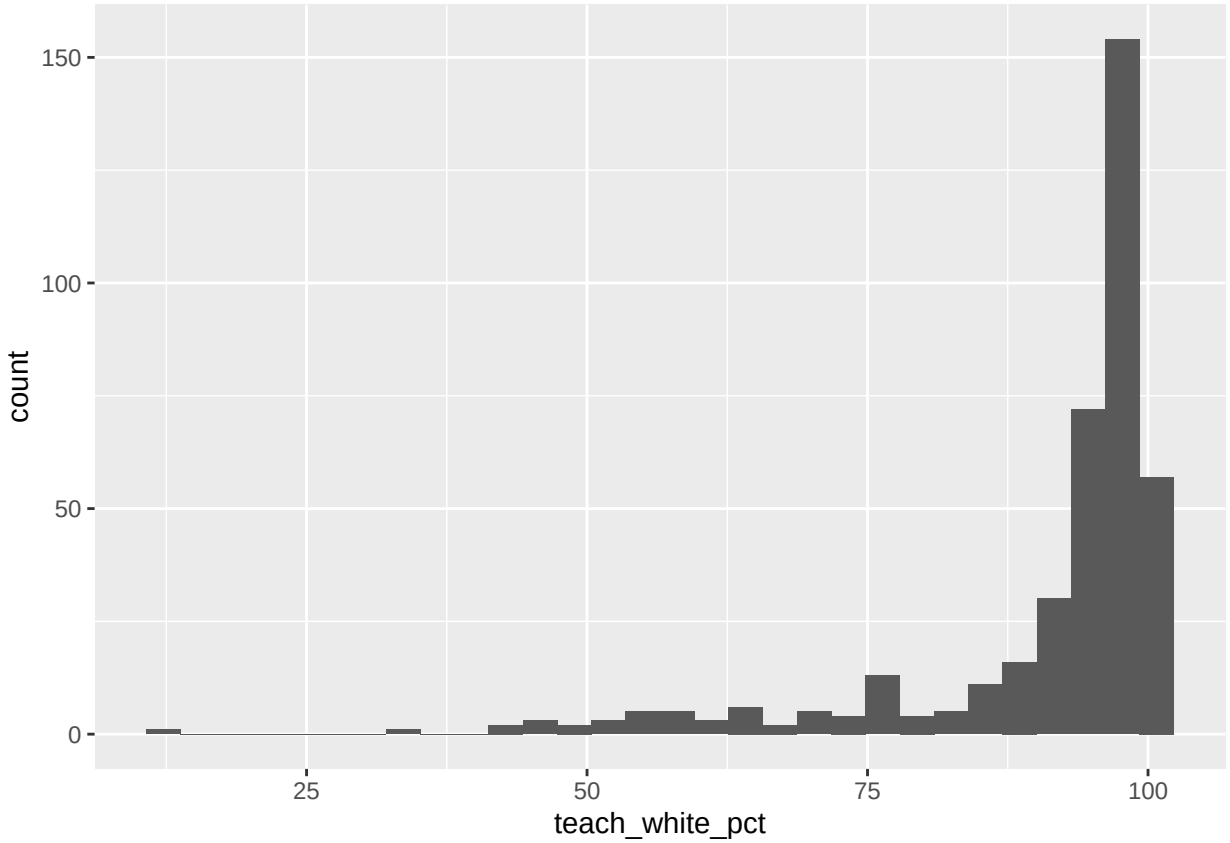


This is bimodal data.

```
tw_plot <- madese %>%
  select(teach_white_pct) %>%
  ggplot(mapping = aes(x = teach_white_pct)) +
  geom_histogram()

tw_plot
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 5 rows containing non-finite values (stat_bin).
```



This data is unimodal and skewed to the left.

```
tw_summary <- madese %>%
  select(teach_white_pct) %>%
  summary()

tw_summary %>% knitr::kable()
```

teach_white_pct
Min. : 11.43
1st Qu.: 91.56
Median : 96.50
Mean : 91.23
3rd Qu.: 98.42
Max. :100.00
NA's :5

```
w_summary <- madese %>%
  select(stud_white_pct) %>%
  summary()

w_summary %>% knitr::kable()
```

stud_white_pct
Min. : 0.30
1st Qu.: 54.50
Median : 79.50
Mean : 67.12
3rd Qu.: 87.80
Max. :100.00
NA's :4

For the student variable, the median proportion is much smaller than the teacher variable. The student variable also has a greater overall spread. The IQR for teachers is less than 10 percentage points.

```
unique(madese$type)
```

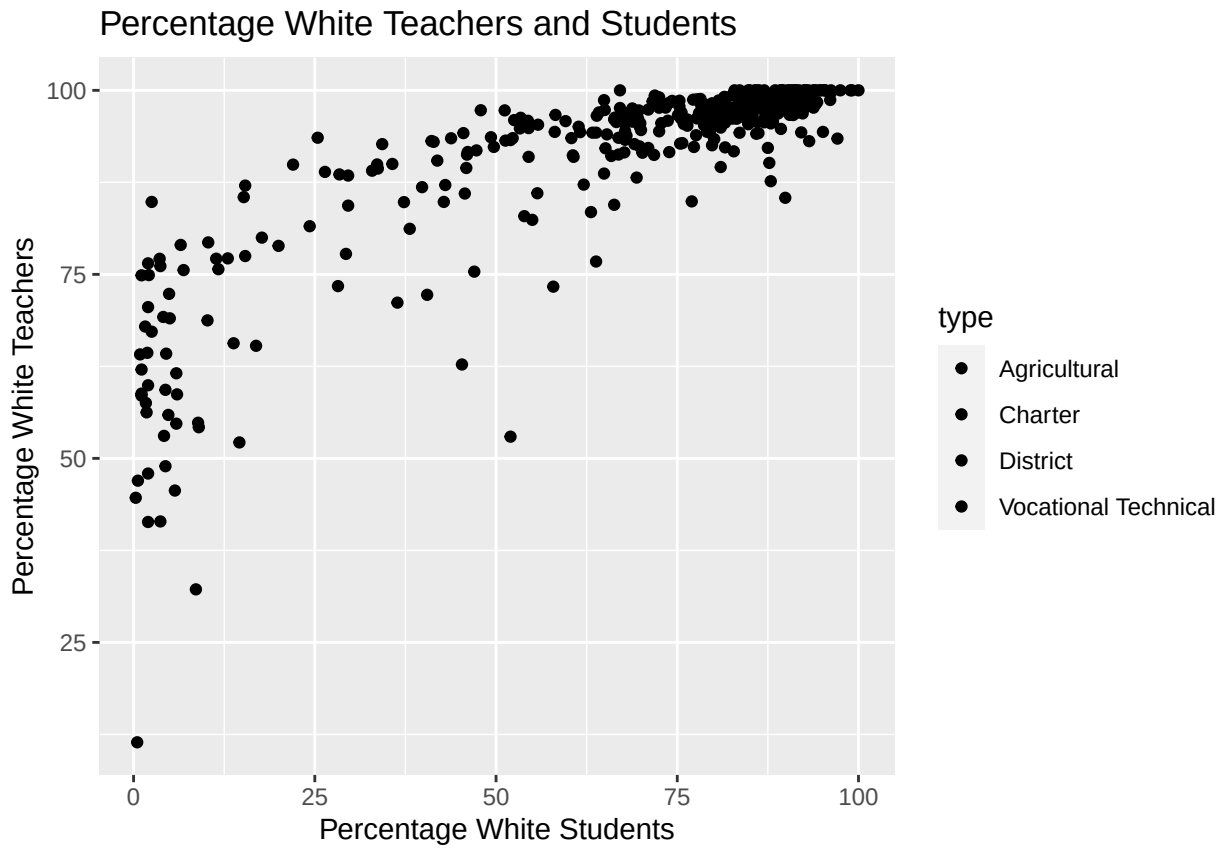
```
## [1] "Charter"              "District"             "Vocational Technical"
## [4] "Agricultural"
```

```
madese_small <- madese %>%
  filter(type %in% c("Charter", "District")) %>%
  group_by(type) %>%
  summarize(percent_w = mean(stud_white_pct, na.rm = TRUE))

knitr::kable(madese_small, digits = 3, col.names = c("School Type", "Percent of White Students"))
```

School Type	Percent of White Students
Charter	30.111
District	76.554

```
madese_scatter <- madese %>%
  ggplot(mapping = aes(x = stud_white_pct, y = teach_white_pct)) +
  geom_point(mapping = aes(fill = type)) +
  scale_color_manual(values = c("lightblue", "pink", "violet", "seagreen")) +
  labs(title = "Percentage White Teachers and Students", x = "Percentage White Students", y = "Percentag

madese_scatter
```

```
## Warning: Removed 5 rows containing missing values (geom_point).
```



According to the data we have a positive exponential direction with most values toward high white teacher and student populations.