Tidytuesday 2021-02-02

Get the Data

```
library(tidyverse)
library(readxl)

tuesdata <- tidytuesdayR::tt_load('2021-02-02')
```

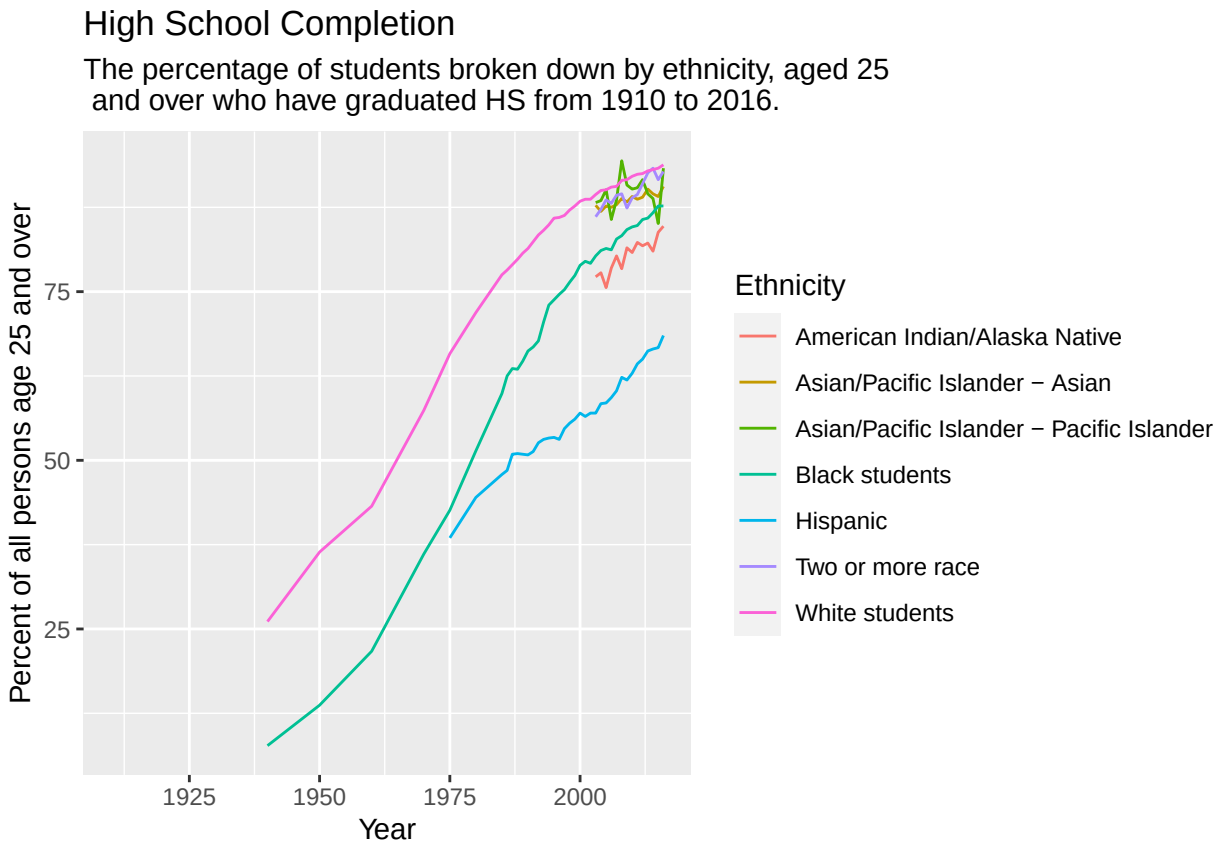
```
##
##  Downloading file 1 of 8: `bach_students.csv`
##  Downloading file 2 of 8: `female_bach_students.csv`
##  Downloading file 3 of 8: `female_hs_students.csv`
##  Downloading file 4 of 8: `hbcu_all.csv`
##  Downloading file 5 of 8: `hbcu_black.csv`
##  Downloading file 6 of 8: `hs_students.csv`
##  Downloading file 7 of 8: `male_bach_students.csv`
##  Downloading file 8 of 8: `male_hs_students.csv`
```

```
hs_students <- tuesdata$hs_students
bach_students <- tuesdata$bach_students
```

Data Wrangling and Visualization

```
hs_students <- hs_students %>%
  mutate(Total = str_sub(Total, 1, 4))

hs_students$Total = as.double(hs_students$Total)

hs_students <- hs_students %>%
  rename(Year = Total)

hs_students <- hs_students %>%
  select(-contains("Standard Errors")) %>%
  select(-contains("Total")) %>%
  pivot_longer(-c(Year), names_to = "Ethnicity", values_to = "Percentage")

hs_students$Percentage = as.double(hs_students$Percentage)

hs_students <- hs_students %>%
  mutate(Ethnicity = case_when(
  Ethnicity == "White1" ~ "White students",
  Ethnicity == "Black1" ~ "Black students",
  Ethnicity == "American Indian/\r\nAlaska Native" ~ "American Indian/Alaska Native",
  TRUE ~ Ethnicity))

hs_students %>%
  ggplot(aes(x = Year, y = Percentage, color = Ethnicity)) +
```

```
  geom_line() +
  labs(x="Year", y="Percent of all persons age 25 and over",
       title = "High School Completion",
       subtitle = "The percentage of students broken down by ethnicity, aged 25\n and over who have grad
```




```
bach_students <- bach_students %>%
  rename(Year = Total)

bach_students <- bach_students %>%
  select(-contains("Standard Errors")) %>%
  select(-contains("Total")) %>%
  pivot_longer(-c(Year), names_to = "Ethnicity", values_to = "Percentage")

bach_students$Percentage = as.double(bach_students$Percentage)

bach_students <- bach_students %>%
  mutate(Ethnicity = case_when(
  Ethnicity == "White1" ~ "White students",
  Ethnicity == "Black1" ~ "Black students",
  Ethnicity == "American Indian/\r\nAlaska Native" ~ "American Indian/Alaska Native",
  TRUE ~ Ethnicity))

bach_students %>%
  ggplot(aes(x = Year, y = Percentage, color = Ethnicity)) +
  geom_line() +
  labs(x="Year", y="Percent of all persons age 25 and over",
       title = "Bachelor's Degree Attainment",
```

```
      subtitle = "The percentage of students broken down by ethnicity, aged 25\n and over who have achi
```