

Week11-scholar

Jiayin Qu

3/28/2020

Library Imports

```
knitr::opts_chunk$set(echo = TRUE)
library(tidyverse)
library(ggplot2)
library(xml2)
library(rvest)
```

Data Import

- imported Dr. Richard Landers Google Scholar page

```
profile_html <- read_html("https://scholar.google.com/citations?user=wMsVHFYAAAAJ&hl=en&oi=ao")

title <- profile_html %>%
  html_nodes(".gsc_a_at") %>%
  html_text()

author <- profile_html %>%
  html_nodes(".gsc_a_at+ .gs_gray") %>%
  html_text()

cite <- profile_html %>%
  html_nodes(".gsc_a_ac") %>%
  html_text()

year <- profile_html %>%
  html_nodes(".gsc_a_hc") %>%
  html_text()

profile_tbl <- as_tibble(cbind(title, author, cite, year)) %>%
  mutate(year = as.numeric(year), cite = as.numeric(cite))
```

Data Import

```
corCite <- cor.test(profile_tbl$year, profile_tbl$cite, method = "pearson")
```

The correlation between year and citation count was -0.3609409 ($p \geq .05$), which is not statistically significant.

Visualization

- visualize the relationship between citation count and year

```
ggplot(profile_tbl, aes(profile_tbl$year, profile_tbl$cite)) +  
  geom_point() +  
  geom_smooth(method = "lm") +  
  labs(x = "year", y = "cite")
```

