

Easy Way #1

Charting graduation and retention rates over time for PSU and some peer institutions.

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse

## v ggplot2 3.2.1      v purrr  0.3.2
## v tibble  2.1.3      v dplyr  0.8.3
## v tidyr   0.8.3      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.4.0

## -- Conflicts ----- tidyverse_conflicts_

## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ipeds)
```

```
## Loading required package: RCurl

## Loading required package: bitops

##
## Attaching package: 'RCurl'

## The following object is masked from 'package:tidyr':
##
##     complete

## Loading required package: Hmisc

## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:dplyr':
##
##     src, summarize

## The following objects are masked from 'package:base':
##
##     format.pval, units

## Loading required package: httr
```

```
library(reshape2)
```

```
##  
## Attaching package: 'reshape2'  
  
## The following object is masked from 'package:tidyr':  
##  
## smiths
```

```
library(ggplot2)  
library(ggthemes)  
library(gridExtra)
```

```
##  
## Attaching package: 'gridExtra'  
  
## The following object is masked from 'package:dplyr':  
##  
## combine
```

```
library(scales)
```

```
##  
## Attaching package: 'scales'  
  
## The following object is masked from 'package:purrr':  
##  
## discard  
  
## The following object is masked from 'package:readr':  
##  
## col_factor
```

```
dir <- "C:\\Users\\mjs26\\Documents\\R\\R-3.6.1\\library\\ipeds\\data\\downloaded"
```

These are the colleges we're interested in, selected by IPEDS ID. This happens to be Portland State University and some of our traditional peer list. You might want to change them...

```
the_colleges <- c(232186, 151111, 122409, 145600, 220862, 228769, 206084, 240453, 172699, 209807 )
```

We pull the most recent IC collection, to get their names and other details:

```
# Load up the IC  
  
directory <- ipeds_survey(table = 'HD', year = 2018, dir=dir)  
names(directory) <- tolower(names(directory))  
  
# Create a unit ID <-> Institution name crosswalk.  
  
getinstnm <- directory$instnm  
names(getinstnm) <- directory$unitid
```

Graduation

Now we get 150% graduation rates. First, we define a function that opens an IPEDS GR for the supplied year, and returns a dataframe containing the adjusted cohort, the number of completers, and the rate.

```
grad_rate_get <- function(year) {  
  gradrates <- (ipeds_survey(table='GR',year=year, dir=dir))  
  names(gradrates) <- tolower(names(gradrates))  
  theRates <- dcast(gradrates, unitid ~ grtype, value.var = 'grtotlt')  
  names(theRates) <- c('unitid','adjusted_cohort','completers')  
  theRates$rate <- theRates$completers/theRates$adjusted_cohort  
  theRates$year <- year  
  return(theRates[, c('unitid','year','adjusted_cohort','completers','rate')])  
}
```

Then we make a list of the years we're interested in.

```
years <- (2012:2018)  
years
```

```
## [1] 2012 2013 2014 2015 2016 2017 2018
```

We run the function for each of those years, and then we combine them into a single (very tall) dataframe. We still have all 7,000+ institutions in the dataset at this point.

```
for (i in seq_along(years)) {  
  if (i==1) {grad_rate <- grad_rate_get(years[i])}  
  else {grad_rate <- rbind(grad_rate, grad_rate_get(years[i]))}  
}
```

Retention

This works just like the graduation rates, and we'll use the same list of years. Note that we're pulling in two different components of the Fall Enrollment collection; one has the rate, and the other has the total enrollment.

```
retention_get <- function(year) {  
  efd <- ipeds_survey(table='EFD', year=year, dir=dir)  
  names(efd) <- tolower(names(efd))  
  efd$retention_rate <- as.numeric(efd$ret.pcf)  
  efd$year <- year  
  return(efd[,c('unitid','year','retention_rate')])  
}  
  
fall_enrollment_get <- function(year) {  
  efa <- ipeds_survey(table='EFA', year=year, dir=dir)  
  names(efa) <- tolower(names(efa))  
  efa <- efa %>%  
    filter(efalevel==2) %>% # This is the total undergraduate enrollment row  
    rename(total_enrollment = eftotlt )  
}
```

```

efa$year <- year
return(efa[,c('unitid', 'year', 'total_enrollment')])
}

for (i in seq_along(years)) {
  if (i==1) {fall_enrollment <- fall_enrollment_get(years[i])}
  else {fall_enrollment <- rbind(fall_enrollment, fall_enrollment_get(years[i]))}
}

for (i in seq_along(years)) {
  if (i==1) {retention <- retention_get(years[i])}
  else {retention <- rbind(retention, retention_get(years[i]))}
}

```

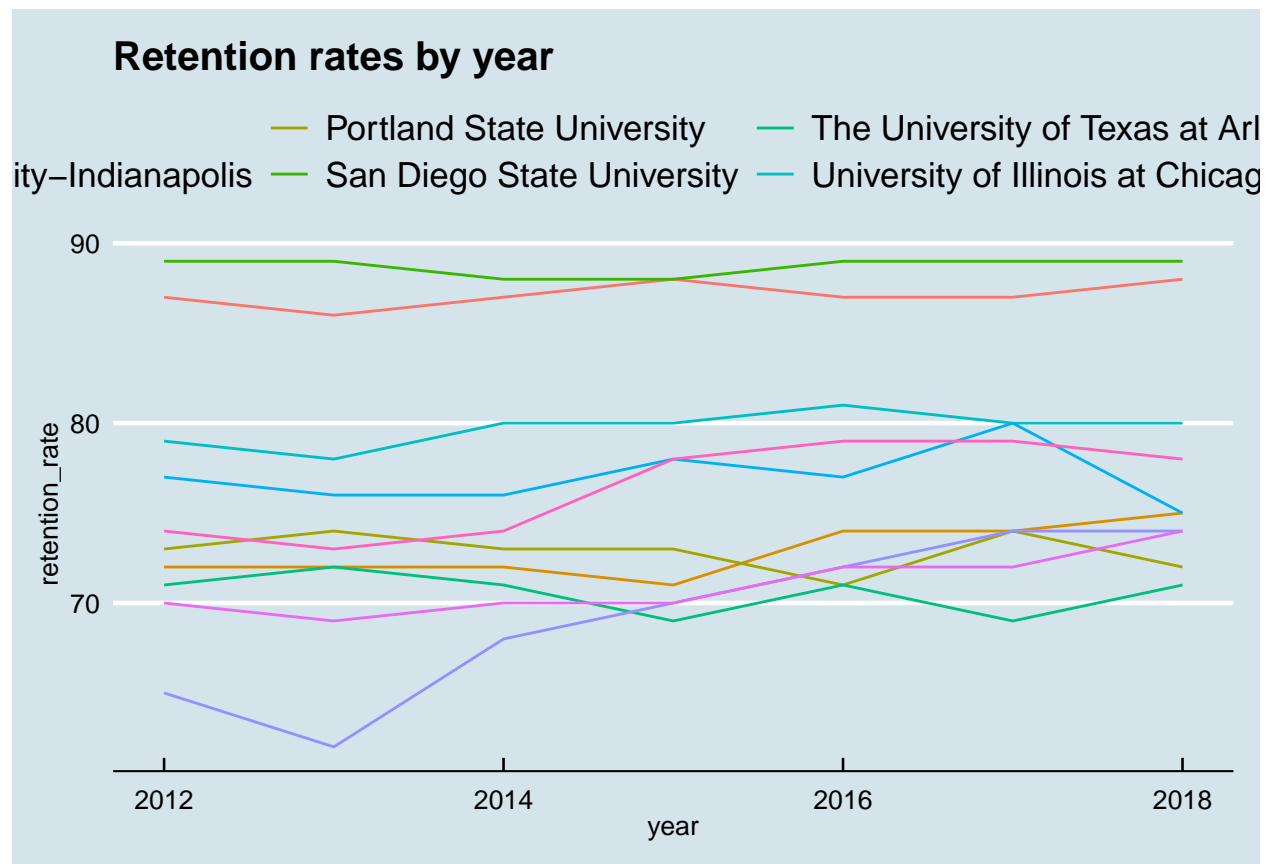
Outputting

I'm just going to use ggplot to do some simple (and mostly bad) charts for illustrative purposes. You might want to build on these and make them better, or export the two datasets from R into your favorite data-viz tool.

```

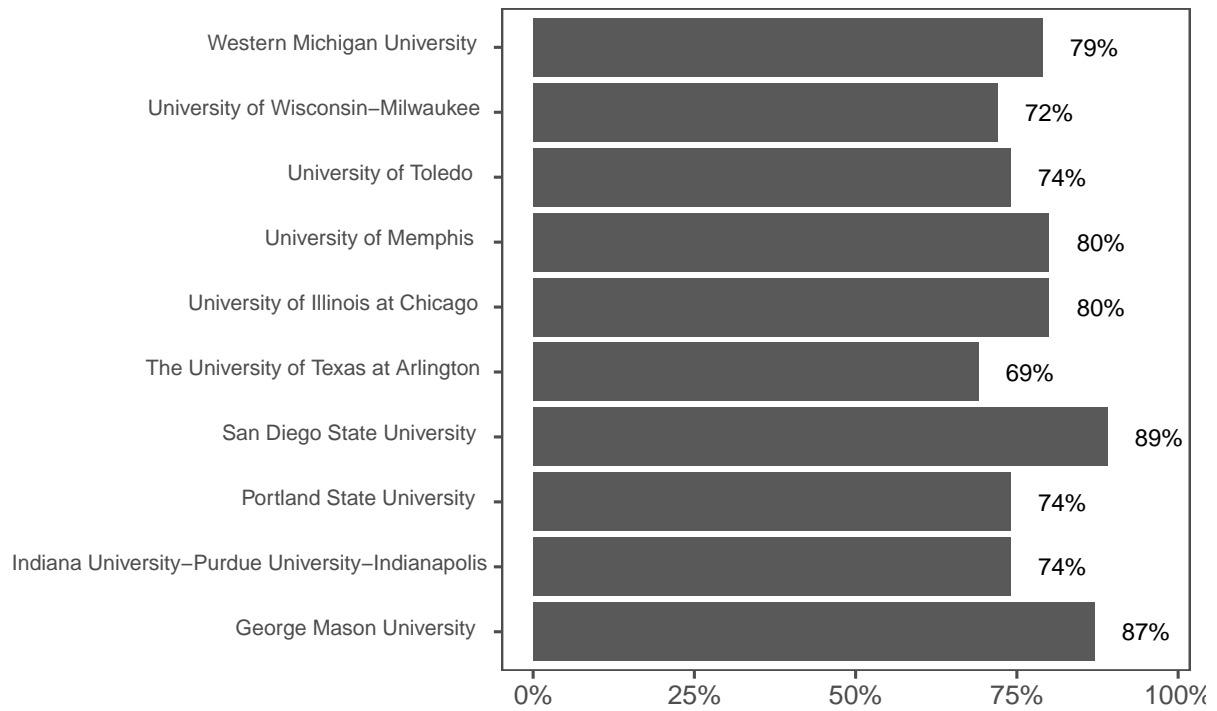
retention %>%
  filter(unitid %in% the_colleges) %>%
  ggplot(aes(year, retention_rate)) +
  theme_economist() +
  geom_line(aes(colour=factor(getinstnm[as.character(unitid)])) ) +
  ggtitle('Retention rates by year')

```

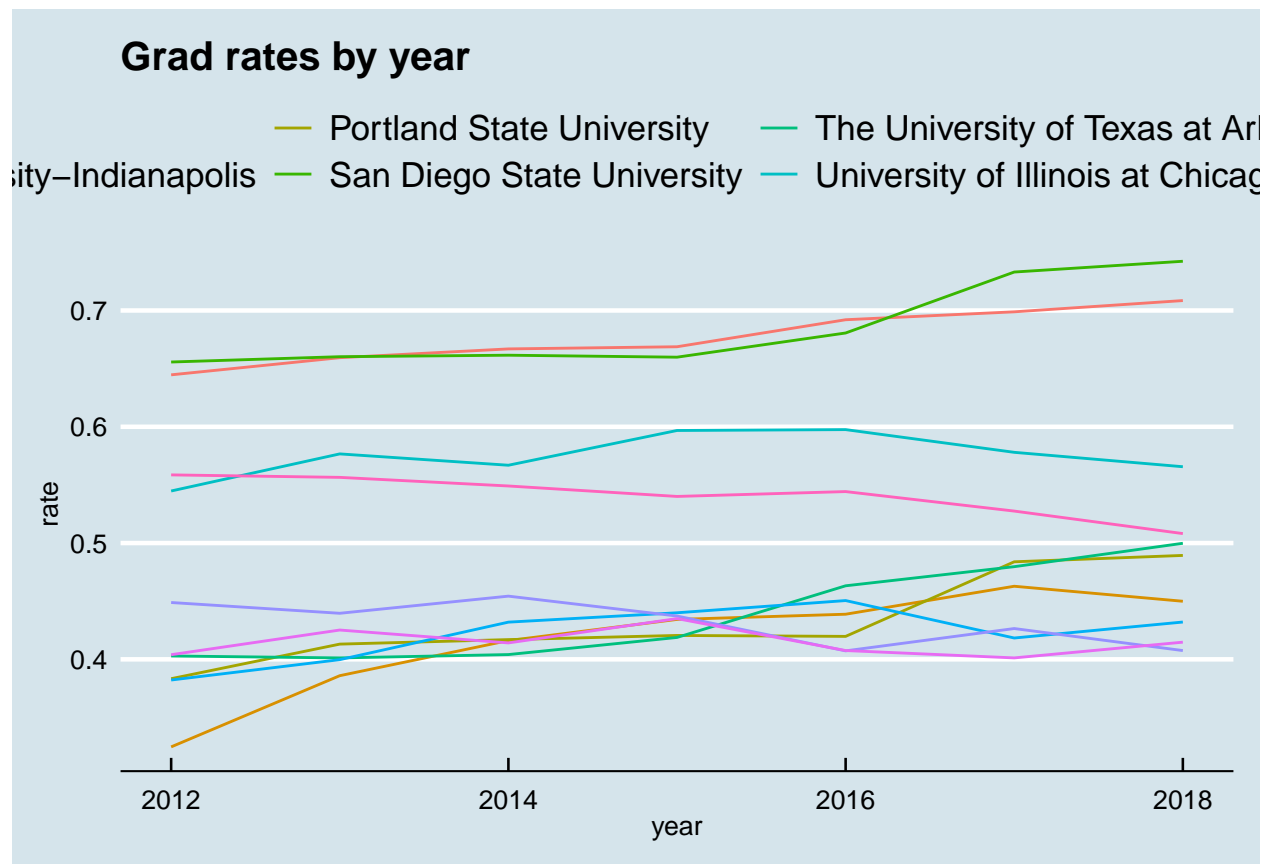


```
retention %>%
  mutate(retention_rate = retention_rate/100) %>%
  filter(unitid %in% the_colleges) %>%
  filter(year == 2017)%>%
  arrange((getinstnm[as.character(unitid)])) %>%
  ggplot(aes((getinstnm[as.character(unitid)]), retention_rate)) +
  geom_bar(stat="identity") +
  theme_few() +
  theme(legend.position="none") +
  scale_y_continuous(labels = scales::percent) +
  geom_text(aes(label=paste0(round(retention_rate*100,1),"%"), y=retention_rate+0.08), size=3) +
  coord_flip() +
  ylab('') +
  xlab('') +
  theme(axis.text.y=element_text(size=8, hjust=0.95,vjust=0.2)) +
  ggtitle('2017 retention rates of PSU and\n peer institutions') +
  theme(plot.title = element_text(hjust = 0.5))
```

2017 retention rates of PSU and peer institutions



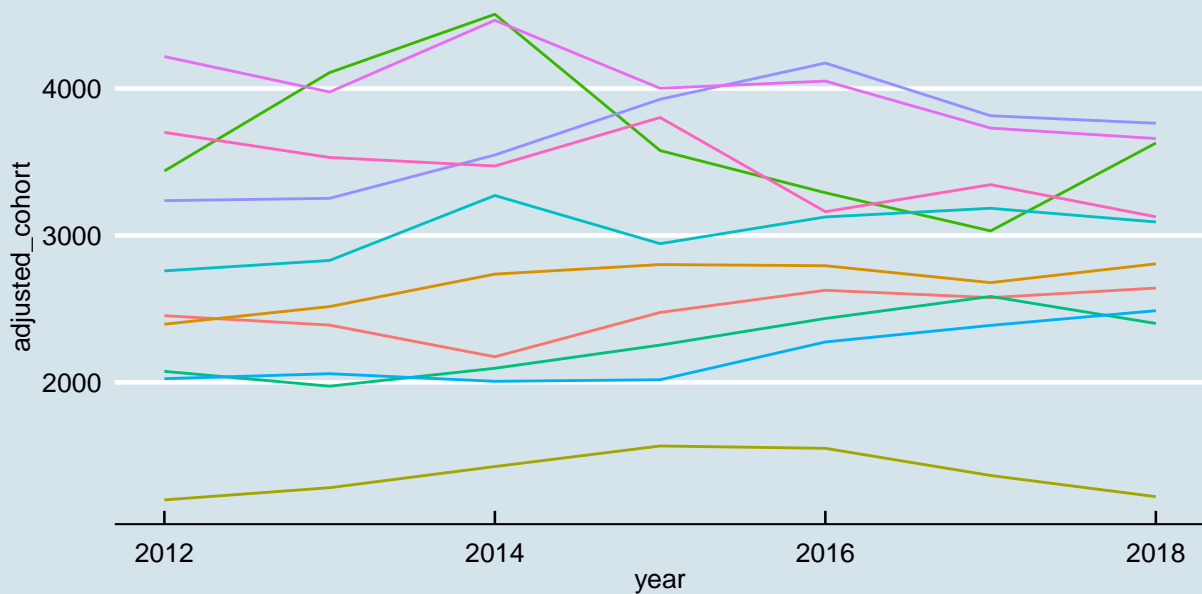
```
grad_rate %>%
  filter(unitid %in% the_colleges) %>%
  ggplot(aes(year, rate)) +
  theme_economist() +
  geom_line(aes(colour=factor(getinstnm[as.character(unitid)])) ) +
  ggtitle('Grad rates by year')
```



```
grad_rate %>%
  filter(unitid %in% the_colleges) %>%
  ggplot(aes(year, adjusted_cohort)) +
  theme_economist() +
  geom_line(aes(colour=factor(getinstnm[as.character(unitid)])) ) +
  ggtitle('Grad cohort size by year')
```

Grad cohort size by year

— Portland State University — The University of Texas at Arlington
— San Diego State University — University of Illinois at Chicago



```

grad_rate %>%
  filter(unitid %in% the_colleges) %>%
  filter(year == 2017)%>%
  arrange((getinstnm[as.character(unitid)])) %>%
  ggplot(aes((getinstnm[as.character(unitid)]), rate)) +
  geom_bar(stat="identity") +
  theme_few() +
  theme(legend.position="none") +
  scale_y_continuous(labels = scales::percent) +
  geom_text(aes(label=paste0(round(rate*100,1),"%"), y=rate+0.08), size=3) +
  coord_flip() +
  ylab('') +
  xlab('') +
  theme(axis.text.y=element_text(size=8, hjust=0.95,vjust=0.2)) +
  ggtitle('2017 six-year graduation rates of PSU and\n peer institutions') +
  theme(plot.title = element_text(hjust = 0.5))
    
```


2017 six-year graduation rates of PSU and peer institutions

