exercise_1_code+output.R

Emmanuel

Wed Mar 27 12:50:12 2019

```
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------- tidyverse
1.2.1 --
```

```
## v ggplot2 3.1.0     v purrr   0.3.0
## v tibble  2.0.1     v dplyr   0.7.8
## v tidyr   0.8.2     v stringr 1.3.1
## v readr   1.3.1     v forcats 0.3.0
```

```
## -- Conflicts ------------------------------------------------
tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(tibble)
library(ggplot2)
library(dplyr)
```

```
cwur <- read.csv("cwurData.csv")
cwurdf <- data.frame(cwur)
```

```
names(cwurdf)
```

```
##  [1] "world_rank"           "institution"          "country"
##  [4] "national_rank"        "quality_of_education" "alumni_employment"
##  [7] "quality_of_faculty"   "publications"         "influence"
## [10] "citations"            "broad_impact"         "patents"
## [13] "score"                "year"
```

```
times <- read.csv("timesData.csv")
timesdf <- data.frame(times)
```

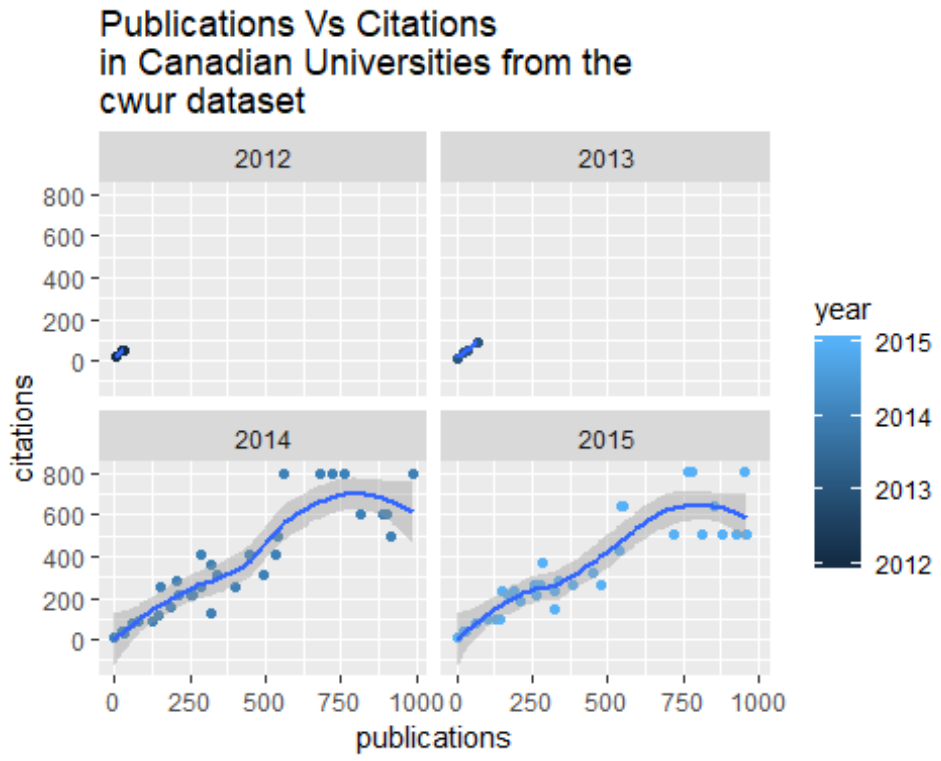
```
names(timesdf)
```

```
##  [1] "world_rank"             "university_name"
##  [3] "country"                "teaching"
##  [5] "international"          "research"
##  [7] "citations"              "income"
##  [9] "total_score"            "num_students"
## [11] "student_staff_ratio"    "international_students"
## [13] "female_male_ratio"      "year"
```

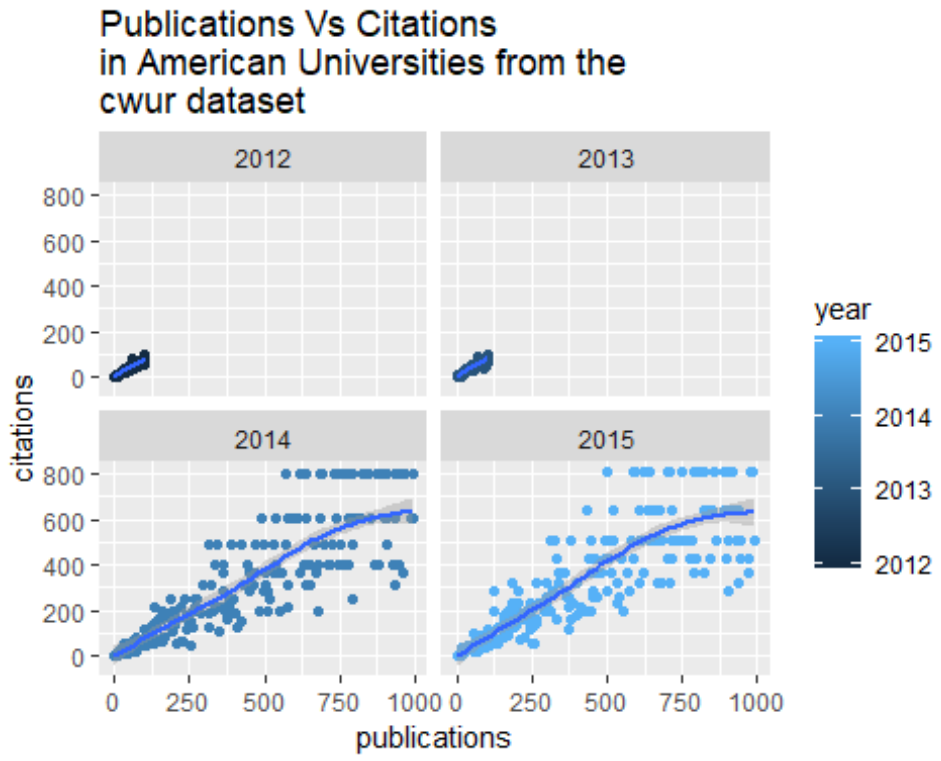
```r
#comparing the research(publications) and citations from both the center for
#world ranking dataset and the times dataset between canadian universities
#and American universities only.

#canadian universities from the cwur dataset
cwurfilter <- cwurdf %>% filter(country == "Canada")
ggplot(cwurfilter, aes(x=publications, y=citations)) + geom_point(aes(color =
year)) +
  geom_smooth(method = "loess", formula = y~x) + facet_wrap(year~.) +
labs(title = "Publications Vs Citations
in Canadian Universities from the
cwur dataset")
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : span too small. fewer data values than degrees of freedom.

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : pseudoinverse used at 6.87

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : neighborhood radius 21.13

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : reciprocal condition number 0

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : There are other near singularities as well. 26.317

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : span too small.
## fewer data values than degrees of freedom.

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : pseudoinverse used
## at 6.87

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : neighborhood
radius
## 21.13

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : reciprocal
## condition number 0

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
```

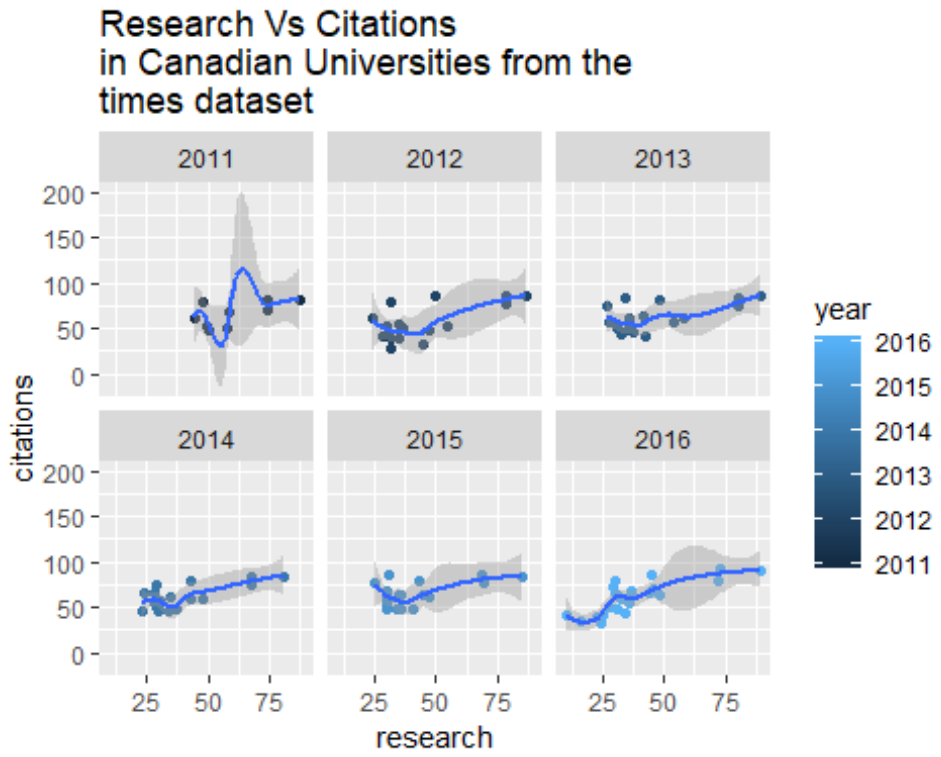
```
## as.matrix(model.frame(delete.response(terms(object)), : There are other
## near singularities as well. 26.317

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : span too small. fewer data values than degrees of freedom.

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : pseudoinverse used at 1.67

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : neighborhood radius 32.33

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : reciprocal condition number 0

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : There are other near singularities as well. 1877.5

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : span too small.
## fewer data values than degrees of freedom.

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : pseudoinverse used
## at 1.67

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : neighborhood
radius
## 32.33

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : reciprocal
## condition number 0

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : There are other
## near singularities as well. 1877.5
```



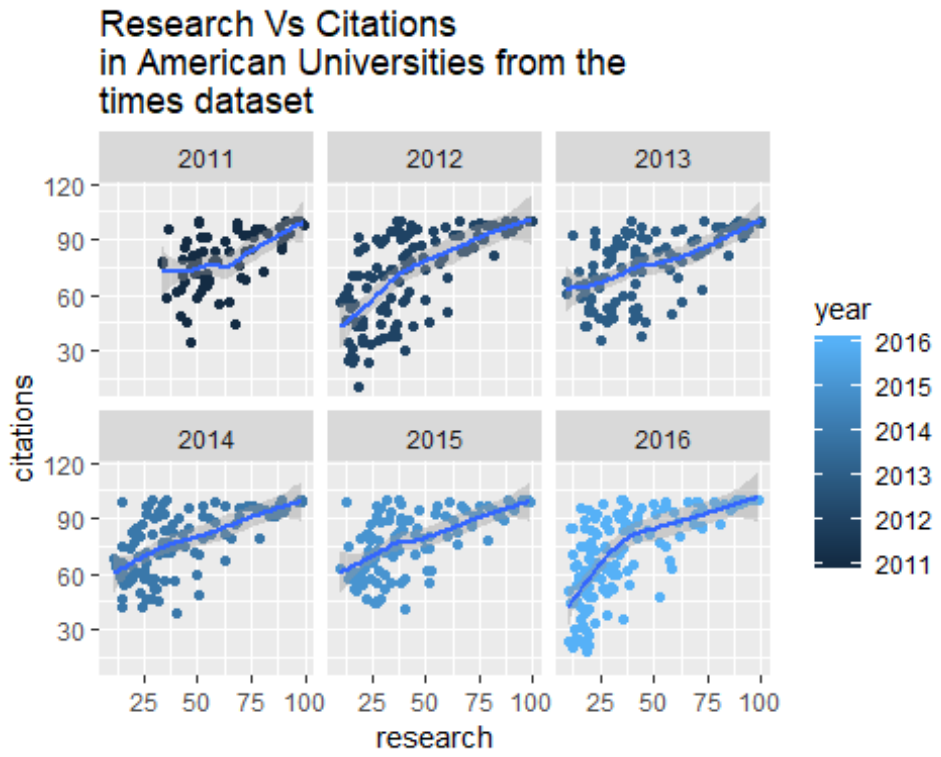

```
#american universities from the cwur dataset
amerfilter <- cwurdf %>% filter(country =="USA")
ggplot(amerfilter, aes(x=publications, y=citations)) + geom_point(aes(color =
year)) +
  geom_smooth(method = "loess", formula = y~x) + facet_wrap(year~.) +
labs(title = "Publications Vs Citations
in American Universities from the
cwur dataset")
```




```r
#canadian universities from the times dataset
timesfilter <- timesdf %>% filter(country == "Canada")
ggplot(timesfilter, aes(x=research, y=citations)) + geom_point(aes(color =
year)) +
  geom_smooth(method = "loess", formula = y~x) + facet_wrap(year~.) +
labs(title = "Research Vs Citations
in Canadian Universities from the
times dataset")
```




```r
#american universities from the times dataset
amertimes <- timesdf %>% filter(country == "United States of America")
ggplot(amertimes, aes(x=research, y=citations)) + geom_point(aes(color =
year)) +
  geom_smooth(method = "loess", formula = y~x) + facet_wrap(year~.) +
labs(title = "Research Vs Citations
in American Universities from the
times dataset")
```


Research Vs Citations
in American Universities from the
times dataset
2011
2012
2013
2014
2015
2016
citations
120
90
60
30
25
50
75
100
research
year
2016
2015
2014
2013
2012
2011