CSSS 512: Lab 1

Logistics & R Refresher

2018-3-30

Agenda

1. **Logistics**
   - Labs, Office Hours, Homeworks
   - Goals and Expectations
   - R, R Studio, R Markdown, LaTeX

2. **Time Series Data in R**
   - Unemployment in Maine
   - Global Temperature
   - Electricity, Beer, and Chocolate Production

3. **Panel Data in R**
   - Democracy and Income
   - Data wrangling

Logistics

1. **Lab Sessions**: Fri, 1:00-2:20pm in Savery 117
   - Covers application of material from lecture using examples; clarification and extention of lecture material; Q & A for homeworks and lectures
   - Materials will be available on the **course website**
2. **Office Hours**: Tues, 3:00-4:20pm in Smith 220
   - Available for trouble shooting and specific questions about homework and lecture materials
3. **Homeworks**: 3-4 due every 2 weeks or so
   - Ideally, done using R or R Studio with write up in LaTeX
   - Using R Studio with R Markdown is an easy way to do this
   - Many packages: `tseries`, `forecast`, `lmtest`, `urca`, `quantmod`, etc.

Logistics

4. When this course is over, you should be able to do the following (and more):
   - Identify and understand time series dynamics: seasonality, deterministic trends, moving average processes, autoregressive processes
   - Distinguish between stationary and nonstationary time series, perform unit root tests, fit ARMA and ARIMA models, use cross validation for model assessment
   - Analyze multiple continuous time series using vector autoregression, perform cointegration tests, and estimate error correction models for cointegrated time series
   - Distinguish between random effects, fixed effects, and mixed effects and decide when each of these are appropriate
   - Understand Nickell bias and use an instrumental variable approach with GMM to address the issue
   - Perform multiple imputation and in-sample simulations for panel data

Logistics

5. The course moves fast: you should comfortable doing the following for the homework assignments and project
   - tidying and transforming data, especially time series and panel data
   - importing and exporting data sets
   - generating plots of your data and results
   - writing basic functions and loops for repeated procedures

- Fortunately, for those of you new to R, there are many resources to get you up to speed
  - Cowpertwait and Metcalfe (2009) - download via UW library
  - Zuur et al. (2009)
  - Wickham and Groleman (2017)

Logistics

6. Please make sure that you have R or R Studio installed on your computer
7. If you would like to learn how to use LaTeX, this is a great opportunity to do so
   - An easy way to get introduced to this is to use R Markdown within R Studio
   - Make sure you have TeX installed, which you can find here
   - Make sure you have R Markdown installed using `install.packages("rmarkdown")`
   - Now in R Studio, choose `File` → `New File` → `R Markdown`

Logistics

8. Using R Markdown
   - Choose to compile your document as a PDF or HTML file and give it a title
   - Now you will be given a template
   - Embed your code within

     ```` ```{r} ````

     and

     ```` ``` ````

     and write up your text outside
   - Then press `Knit` and it will produce a PDF or HTML document with your code, R output, and text nicely formatted
   - Please try to complete your homeworks in this way

Questions

Time Series Data - Unemployment in Maine

```r
# Acquire the data
#   Monthly unemployment in Maine from January 1996 to August 2006
www <- "http://students.washington.edu/dhyoo/Maine.dat"
Maine.month <- read.table(www, header = TRUE)

# Attach the object and check its class
attach(Maine.month)
class(Maine.month)
```

```
## [1] "data.frame"
```

```r
#Monthly unemployment data
head(Maine.month)
```

```
##   unemploy
## 1      6.7
## 2      6.7
## 3      6.4
## 4      5.9
## 5      5.2
## 6      4.8
```

Time Series Data - Unemployment in Maine

```
# Create a time series object
help(ts)
Maine.month.ts <- ts(unemploy, start = c(1996, 1), freq = 12)
Maine.month.ts
```

```
##      Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
## 1996 6.7 6.7 6.4 5.9 5.2 4.8 4.8 4.0 4.2 4.4 5.0 5.0
## 1997 6.4 6.5 6.3 5.9 4.9 4.8 4.5 4.0 4.1 4.3 4.8 5.0
## 1998 6.2 5.7 5.6 4.6 4.0 4.2 4.1 3.6 3.7 4.1 4.3 4.0
## 1999 4.9 5.0 4.6 4.3 3.9 4.0 3.6 3.3 3.1 3.3 3.7 3.7
## 2000 4.4 4.4 4.1 3.5 3.1 3.0 2.8 2.5 2.6 2.8 3.1 3.0
## 2001 3.9 4.2 4.0 4.1 3.5 3.5 3.4 3.1 3.4 3.7 4.0 4.0
## 2002 5.0 4.9 5.0 4.7 4.0 4.2 4.0 3.6 3.7 3.9 4.5 4.6
## 2003 5.6 5.8 5.6 5.5 4.8 4.9 4.8 4.3 4.5 4.6 4.8 4.7
## 2004 5.6 5.6 5.5 4.8 4.2 4.3 4.2 3.8 4.0 4.2 4.6 4.6
## 2005 5.5 5.8 5.5 5.2 4.7 4.6 4.5 4.1 4.4 4.4 4.8 4.6
## 2006 5.3 5.6 4.9 4.6 4.2 4.4 4.4 3.9
```

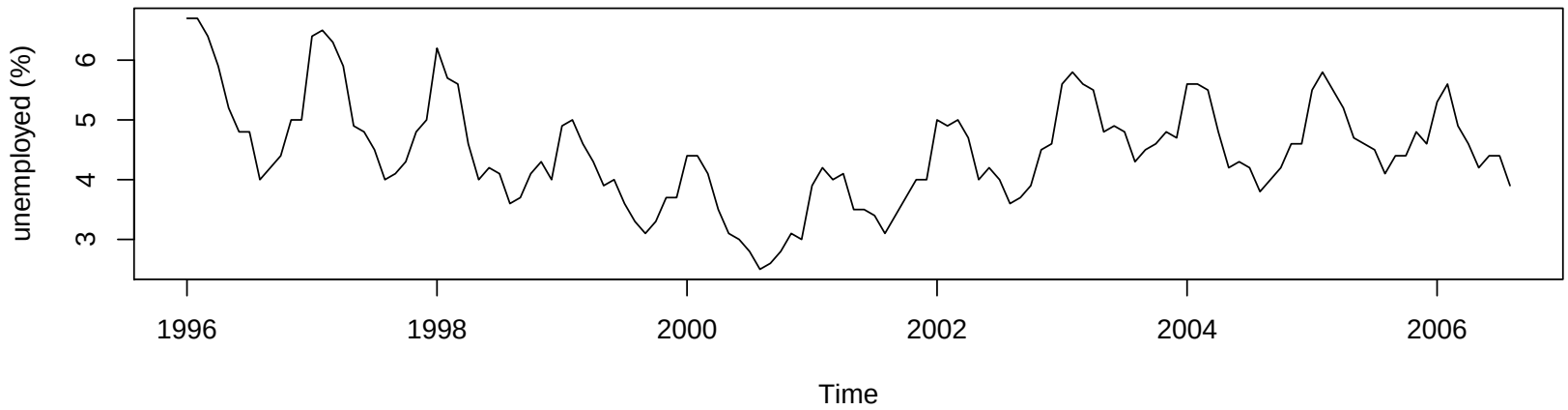
Time Series Data - Unemployment in Maine

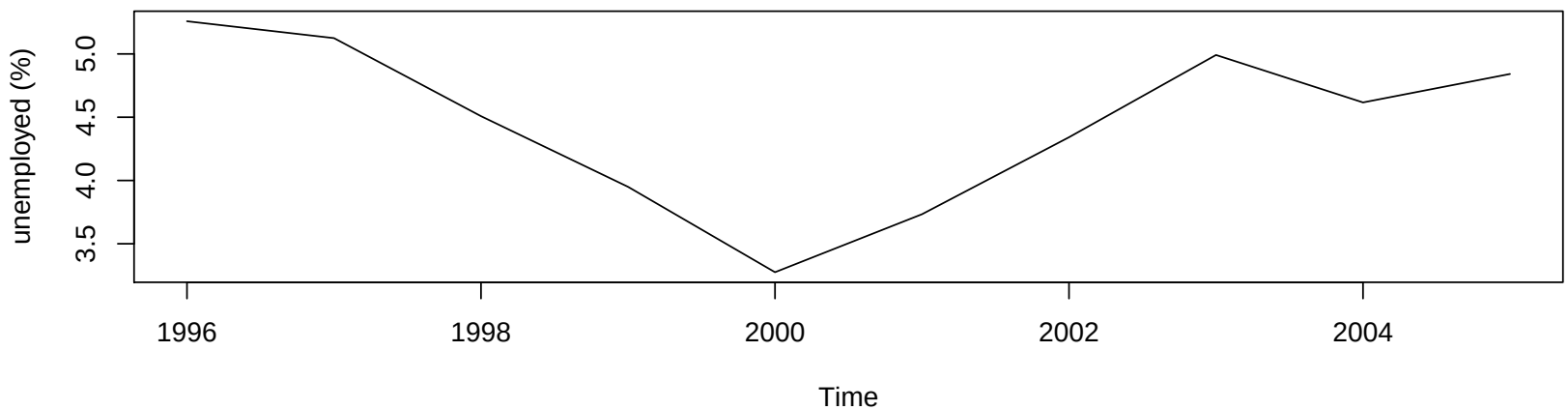
```
# Find the mean unemployment per year
Maine.annual.ts <- aggregate(Maine.month.ts)/12
Maine.annual.ts
```

```
## Time Series:
## Start = 1996
## End = 2005
## Frequency = 1
##  [1] 5.258333 5.125000 4.508333 3.950000 3.275000 3.733333 4.341667
##  [8] 4.991667 4.616667 4.841667
```

Time Series Data - Unemployment in Maine

```
# Plot the time series. Intuitively, how would you describe the pattern of unemployment?
layout(1:2)
plot(Maine.month.ts, ylab="unemployed (%)")
plot(Maine.annual.ts, ylab="unemployed (%)")
```





Time Series Data - Unemployment in Maine

```r
# Find unmployment rates for Feburary and August
Maine.Feb <- window(Maine.month.ts, start = c(1996,2), freq = TRUE)
Maine.Aug <- window(Maine.month.ts, start = c(1996,8), freq = TRUE)
# Find ratio of mean unemployment in Feb and August versus grand mean
Feb.ratio <- mean(Maine.Feb) / mean(Maine.month.ts)
Aug.ratio <- mean(Maine.Aug) / mean(Maine.month.ts)

Maine.Feb
```

```
## Time Series:
## Start = 1996.083
## End = 2006.083
## Frequency = 1
##  [1] 6.7 6.5 5.7 5.0 4.4 4.2 4.9 5.8 5.6 5.8 5.6
```

```r
Feb.ratio
```

```
## [1] 1.222529
```

```r
Aug.ratio
```

```
## [1] 0.8163732
```

Time Series Data - Global Temperature

```r
# Acquire the data
www <- "http://students.washington.edu/dhyoo/global.dat"
# Average global temperature from Univ. East Anglia and UK Met Office
# Monthly from January 1856 to December 2005
Global <- scan(www)
```

1. Create a time series object using the data that starts in Jan 1856 and ends in Dec 2005 with monthly observations.

2. Find the mean temperature for each year and save in a new time series object.

3. Plot the two objects.

4. Observe global temperature from 1970 to 2005 using the window function and plot.

Time Series Data - Global Temperature

```r
# Create a time series object
Global.ts <- ts(Global, st = c(1856, 1), end = c(2005, 12), fr = 12)
head(Global.ts)
```

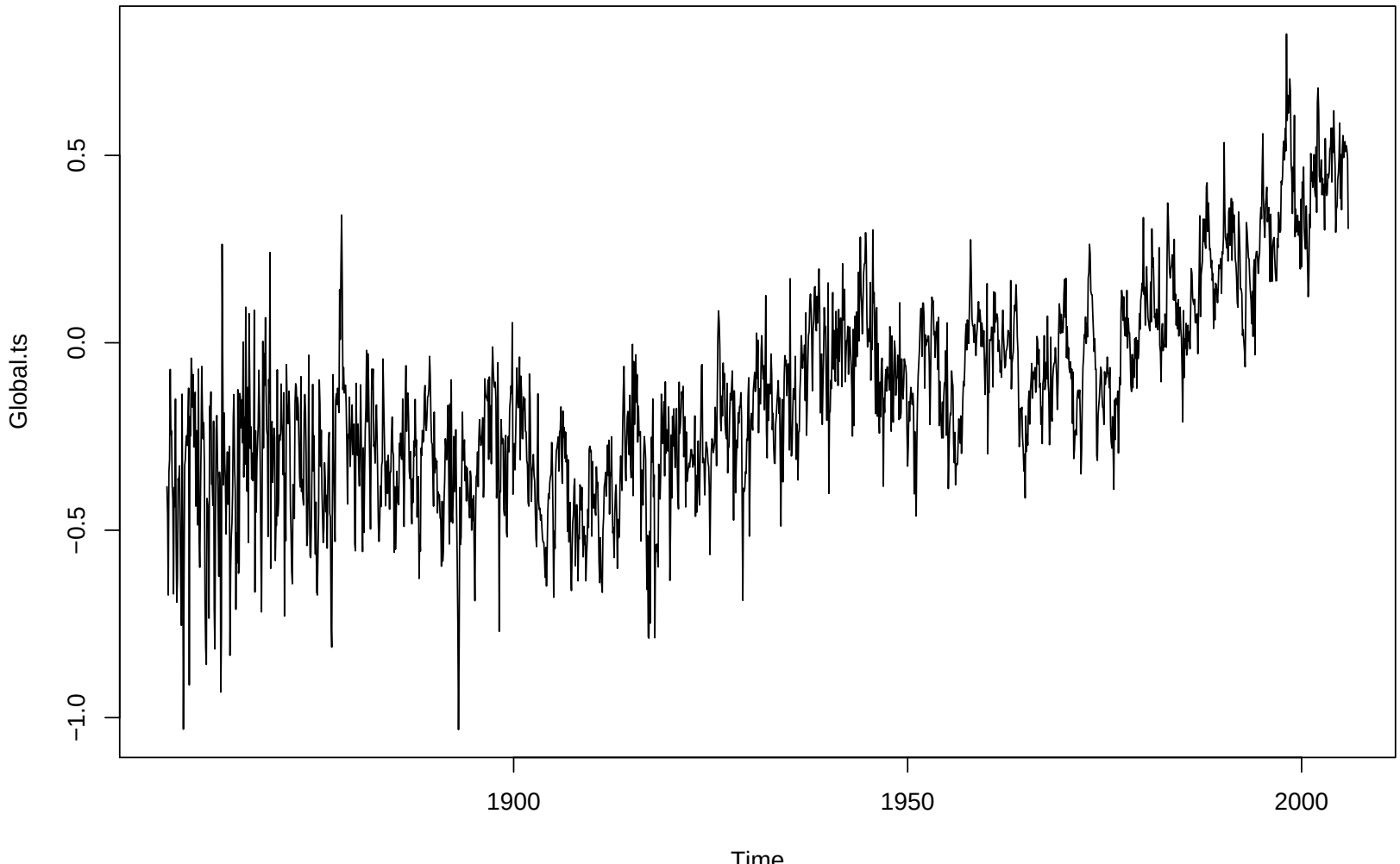
```
## [1] -0.384 -0.457 -0.673 -0.344 -0.311 -0.071
```

```r
# Find the mean temperature for each year
Global.annual <- aggregate(Global.ts, FUN = mean)
head(Global.annual)
```

```
## [1] -0.3812500 -0.4611667 -0.4153333 -0.2252500 -0.3697500 -0.4003333
```

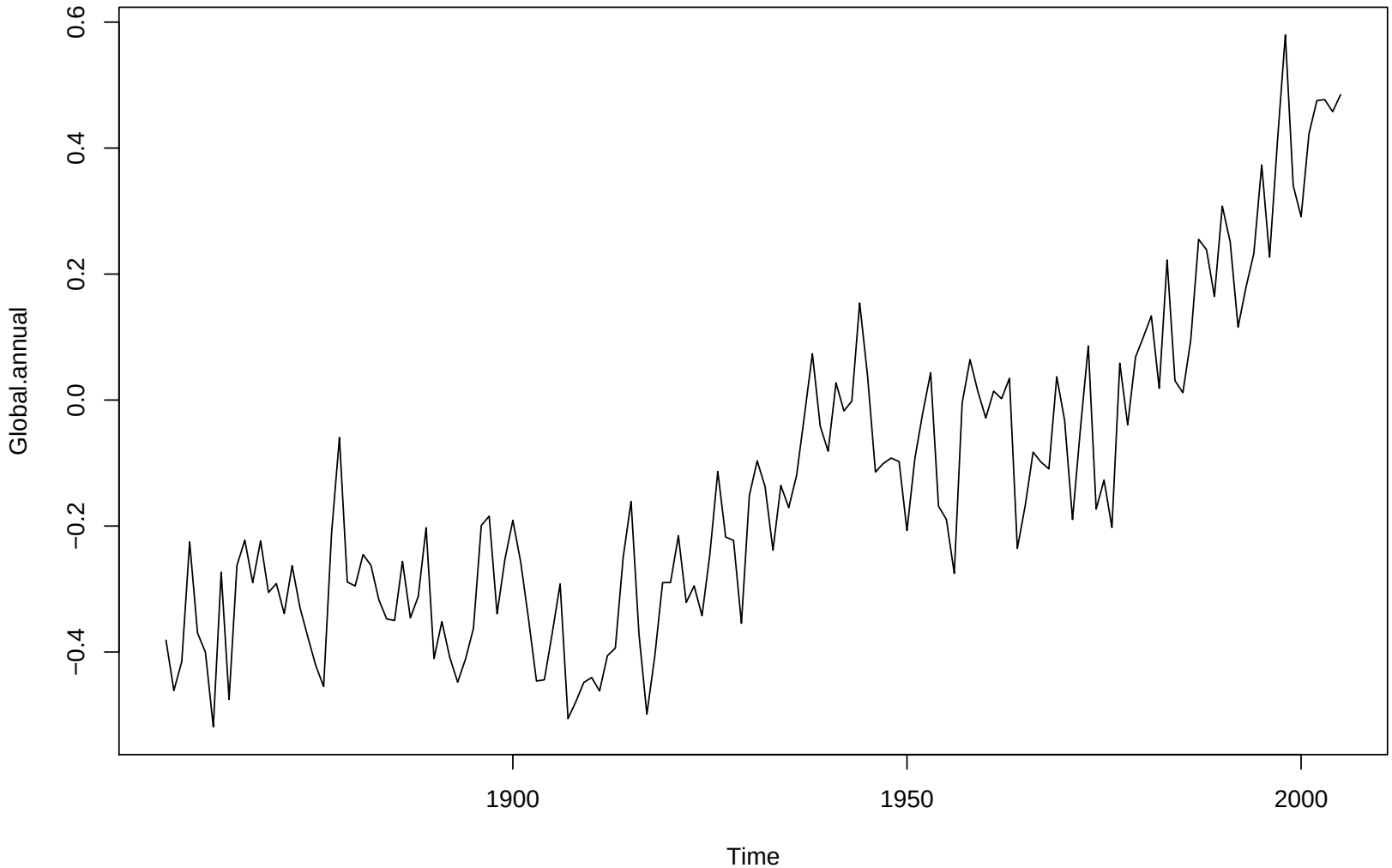
Time Series Data - Global Temperature

```
# Plot the time series.
# How would you describe the pattern in global temperature?
plot(Global.ts)
```



Time Series Data - Global Temperature

```
plot(Global.annual)
```



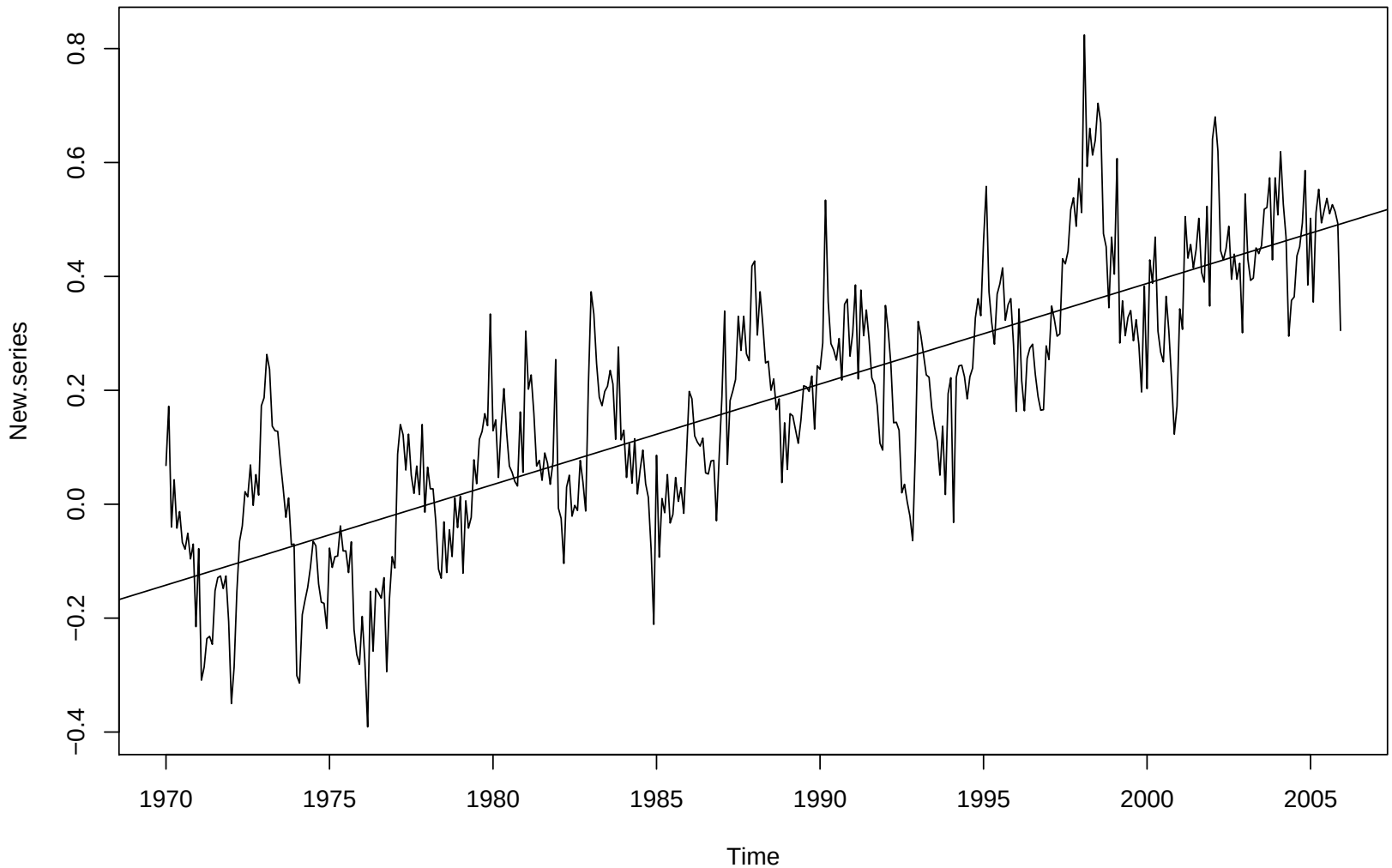
Time Series Data - Global Temperature

```
# Observe between 1970 and 2005 only
New.series <- window(Global.ts, start=c(1970, 1), end=c(2005, 12))

# Express each month fractionally
New.time <- time(New.series)
```

Time Series Data - Global Temperature

```
# How would you describe this pattern?
plot(New.series); abline(reg=lm(New.series ~ New.time))
```



Multiple Time Series - Electricty, Beer, Chocolate Production

```
# Acquire the data
www <- "http://students.washington.edu/dhyoo/cbe.dat"
#   Electricity (millions of kWh), beer (Ml), and chocolate production (tonnes)
# in Australia from January 1958 to December 1990
# from the Australian Bureau of Statistics

CBE <- read.table(www, header=T)

CBE[1:4,]
```

```
##   choc beer elec
## 1 1451 96.3 1497
## 2 2037 84.4 1463
## 3 2477 91.2 1648
## 4 2785 81.9 1595
```

```
class(CBE)
```

```
## [1] "data.frame"
```

Multiple Time Series - Electricty, Beer, Chocolate Production

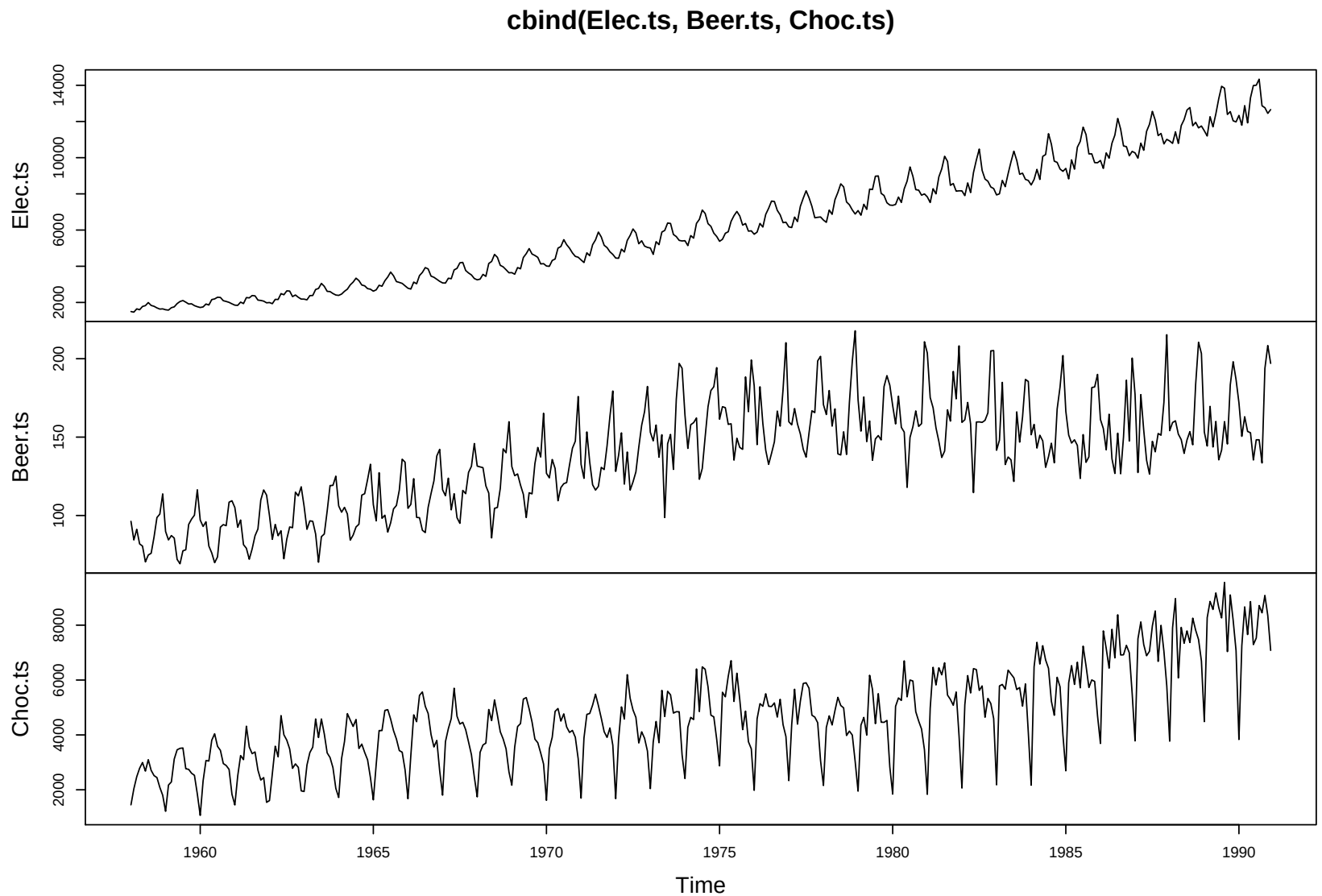
```
# Create separate time series objects for each
Elec.ts <- ts(CBE[, 3], start = 1958, freq = 12)

Beer.ts <- ts(CBE[, 2], start = 1958, freq = 12)

Choc.ts <- ts(CBE[, 1], start = 1958, freq = 12)
```

Multiple Time Series - Electricty, Beer, Chocolate Production

```
plot(cbind(Elec.ts, Beer.ts, Choc.ts))
```



Panel Data - Democracy and Income

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

```
## Loading tidyverse: ggplot2
## Loading tidyverse: tibble
## Loading tidyverse: tidyr
## Loading tidyverse: readr
## Loading tidyverse: purrr
## Loading tidyverse: dplyr
```

```
## Conflicts with tidy packages ----------------------------------------------
```

```
## filter(): dplyr, stats
## lag():    dplyr, stats
```

```
library(ggplot2)

setwd("/Users/danielyoo/CSSS-POLS-512/Labs")
data<-read.csv("Lab1data.csv", header=T)
#Democracy and income data from 174 countries from 2000 to 2010
```

Panel Data - Democracy and Income

```
head(unique(data$country)) # observations on 174 countries
```

```
## [1] Antigua and Barbuda Afghanistan         Albania
## [4] Algeria             Andorra             Angola
## 174 Levels: Afghanistan Albania Algeria Andorra ... Zimbabwe
```

```
head(tapply(data$country, data$Year, length))
```

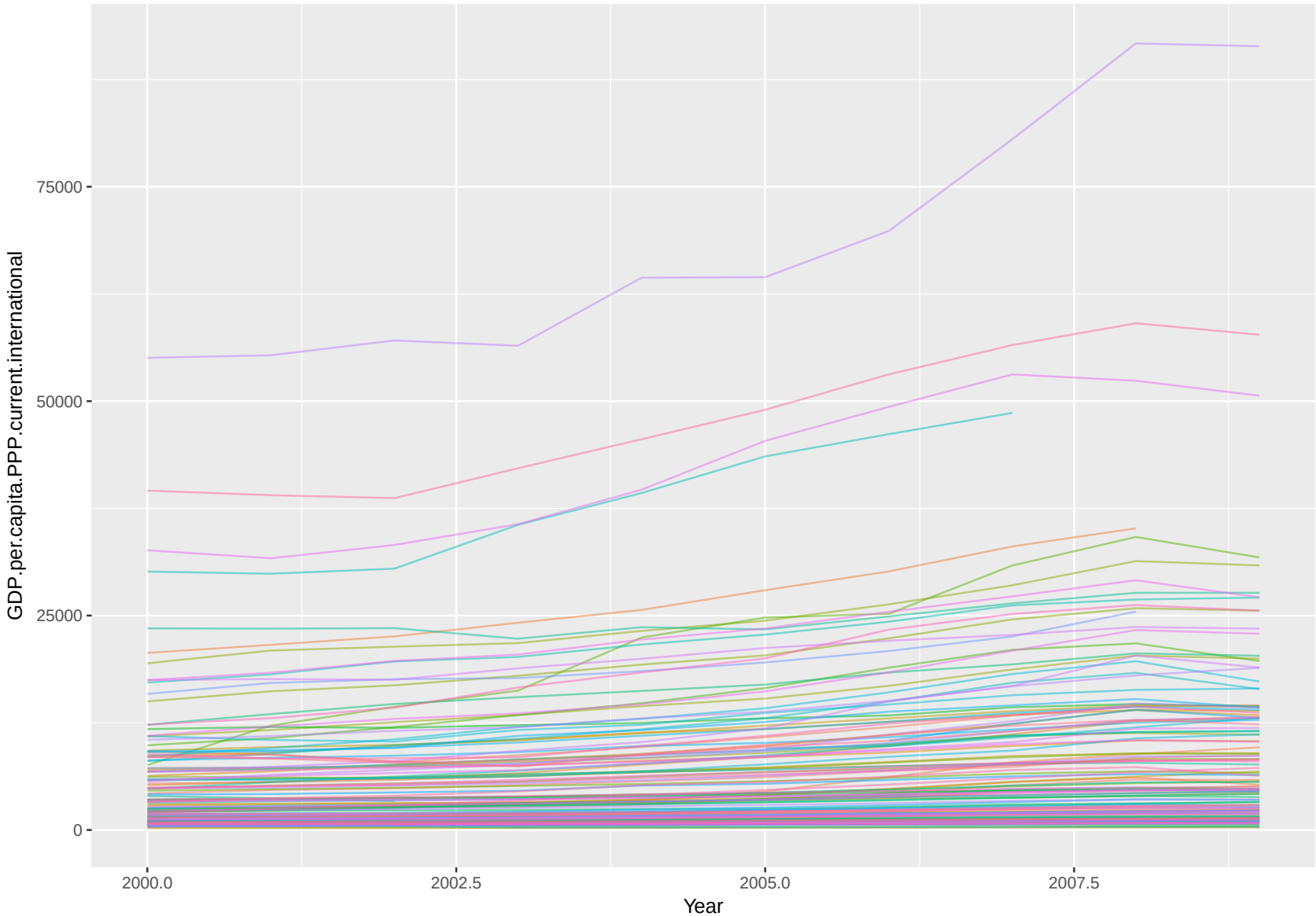
```
## 2000 2001 2002 2003 2004 2005
##  174  174  174  174  174  174
```

```
head(tapply(data$Year, data$country, length))
```

```
##         Afghanistan             Albania             Algeria
##                  11                  11                  11
##             Andorra              Angola Antigua and Barbuda
##                  11                  11                  11
```

Panel Data - Democracy and Income

```
p <- ggplot(data = na.omit(data), aes(x = Year, y = GDP.per.capita.PPP.current.international,
                                      group=country, color=country))
p + geom_line(alpha=0.5) + guides(color=FALSE)
```



Panel Data - Democracy and Income

Some wrangling exercises:

1. Subset the data frame to show only country name and GDP per capita
2. Rearrange the columns of the data frame ascending by polity score
3. Show only values of GDP per capita for South Africa from 2002 to 2008
4. Create a new variable that takes the first letter of the country and attaches it to the year of observation
5. Find the mean of GDP per capita for each year of observation

Panel Data - Democracy and Income

```
library(tidyverse)
head(select(data, country, GDP.per.capita.PPP.current.international))
```

```
##               country GDP.per.capita.PPP.current.international
## 1 Antigua and Barbuda                                  12345.82
## 2 Antigua and Barbuda                                  12654.92
## 3 Antigua and Barbuda                                  12959.93
## 4 Antigua and Barbuda                                  13699.04
## 5 Antigua and Barbuda                                  14866.37
## 6 Antigua and Barbuda                                  15791.64
```

```
head(data[, c(1,3)])
```

```
##               country GDP.per.capita.PPP.current.international
## 1 Antigua and Barbuda                                  12345.82
## 2 Antigua and Barbuda                                  12654.92
## 3 Antigua and Barbuda                                  12959.93
## 4 Antigua and Barbuda                                  13699.04
## 5 Antigua and Barbuda                                  14866.37
## 6 Antigua and Barbuda                                  15791.64
```

```
head(data.frame(data$country, data$GDP.per.capita.PPP.current.international))
```

```
##          data.country data.GDP.per.capita.PPP.current.international
## 1 Antigua and Barbuda                                       12345.82
## 2 Antigua and Barbuda                                       12654.92
## 3 Antigua and Barbuda                                       12959.93
## 4 Antigua and Barbuda                                       13699.04
## 5 Antigua and Barbuda                                       14866.37
## 6 Antigua and Barbuda                                       15791.64
```

Panel Data - Democracy and Income

```
head(arrange(data, polity2))
```

```
##   country Year GDP.per.capita.PPP.current.international polity2
## 1  Bhutan 2000                                2436.943     -10
## 2  Bhutan 2001                                2587.442     -10
## 3  Bhutan 2002                                2775.398     -10
## 4  Bhutan 2003                                2984.397     -10
## 5  Bhutan 2004                                3219.421     -10
## 6   Qatar 2000                               55053.515     -10
```

```
head(data[order(data$polity2),])
```

```
##      country Year GDP.per.capita.PPP.current.international polity2
## 166   Bhutan 2000                                2436.943     -10
## 167   Bhutan 2001                                2587.442     -10
## 168   Bhutan 2002                                2775.398     -10
## 169   Bhutan 2003                                2984.397     -10
## 170   Bhutan 2004                                3219.421     -10
## 1387   Qatar 2000                               55053.515     -10
```

Panel Data - Democracy and Income

```
head(filter(data, country==c("South Africa"), Year>=2002 & Year<=2008))
```

```
##         country Year GDP.per.capita.PPP.current.international polity2
## 1 South Africa 2002                                  7244.218       9
## 2 South Africa 2003                                  7522.254       9
## 3 South Africa 2004                                  7992.767       9
## 4 South Africa 2005                                  8596.831       9
## 5 South Africa 2006                                  9269.283       9
## 6 South Africa 2007                                 10002.543       9
```

```
head(subset(data, data$country==c("South Africa") & data$Year>=2002 & Year<=2008))
```

```
##            country Year GDP.per.capita.PPP.current.international polity2
## 1444 South Africa 2002                                  7244.218       9
## 1445 South Africa 2003                                  7522.254       9
## 1446 South Africa 2004                                  7992.767       9
## 1447 South Africa 2005                                  8596.831       9
## 1448 South Africa 2006                                  9269.283       9
## 1449 South Africa 2007                                 10002.543       9
```

Panel Data - Democracy and Income

```
head(mutate(data, paste(substring(data$country, 1, 1), data$Year, sep="")))
```

```
##               country Year GDP.per.capita.PPP.current.international
## 1 Antigua and Barbuda 2000                                  12345.82
## 2 Antigua and Barbuda 2001                                  12654.92
## 3 Antigua and Barbuda 2002                                  12959.93
## 4 Antigua and Barbuda 2003                                  13699.04
## 5 Antigua and Barbuda 2004                                  14866.37
## 6 Antigua and Barbuda 2005                                  15791.64
##   polity2 paste(substring(data$country, 1, 1), data$Year, sep = "")
## 1      NA                                                    A2000
## 2      NA                                                    A2001
## 3      NA                                                    A2002
## 4      NA                                                    A2003
## 5      NA                                                    A2004
## 6      NA                                                    A2005
```

Panel Data - Democracy and Income

```
data%>%
  group_by(Year)%>%
  summarize(mean(GDP.per.capita.PPP.current.international, na.rm=T)
  )
```

```
## # A tibble: 11 x 2
##     Year `mean(GDP.per.capita.PPP.current.international, na.rm = T)`
##    <int>                                                       <dbl>
##  1  2000                                                    6184.662
##  2  2001                                                    6358.006
##  3  2002                                                    6560.959
##  4  2003                                                    6914.076
##  5  2004                                                    7493.982
##  6  2005                                                    8111.782
##  7  2006                                                    8760.581
##  8  2007                                                    9552.053
##  9  2008                                                    9632.327
## 10  2009                                                    9212.808
## 11  2010                                                         NaN
```