hci+d lab.

Week 7 Extra: Time-Series Graph in R

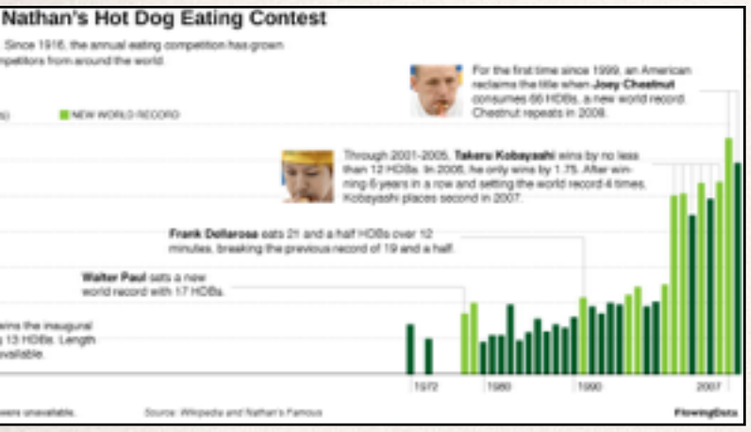
정보 비주얼라이제이션 2015 Fall

human-computer interaction + design lab.
Joonhwan Lee

Time-Series Graph 만들기

Lab 1: Bar Chart

- R을 이용해 Bar Chart 만들기
  - 핫도그 먹기대회 우승자 데이터 hot-dog-contest-winners.csv
    - 매년 미국 독립기념일인 7월 4일에 열리는 행사로 15분 안에 10-20개의 핫도그를 먹는다. 그러나 2001년 일본의 다케루 고바야시가 등장하면서 우승을 위해서는 적어도 50개의 핫도그를 먹어야...
  - 위키피디아 데이터를 csv로 변환
    - http://en.wikipedia.org/wiki/Nathan's_Hot_Dog_Eating_Contest



데이터 샘플

```
"Year","Winner","Dogs eaten","Country","New record"
1980,"Paul Siederman & Joe Baldini",9.1,"United States",0
1981,"Thomas DeBerry ",11,"United States",0
1982,"Steven Abrams ",11,"United States",0
1983,"Luis Llamas ",19.5,"Mexico",0
1984,"Birgit Felden ",9.5,"Germany",0
1985,"Oscar Rodriguez ",11.75,"United States",0
1986,"Mark Heller ",15.5,"United States",0
1987,"Don Wolfman ",12,"United States",0
```

Lab 1: Bar Chart

- 데이터 불러오기
  - `hotdogs <- read.csv("hot-dog-contest-winners.csv", sep=",", header=TRUE);`
    - `read.csv: read csv file format`

```
> hotdogs
   Year                    Winner Dogs.eaten       Country New.record
1  1980 Paul Siederman & Joe Baldini       9.10 United States          0
2  1981               Thomas DeBerry      11.00 United States          0
3  1982                Steven Abrams      11.00 United States          0
4  1983                  Luis Llamas      19.50        Mexico          0
5  1984                Birgit Felden       9.50       Germany          0
6  1985              Oscar Rodriguez      11.75 United States          0
7  1986                 Mark Heller      15.50 United States          0
8  1987                 Don Wolfman      12.00 United States          0
9  1988                   Jay Green      14.00 United States          0
10 1989                   Jay Green      13.00 United States          0
11 1990                 Mike DeVito      16.00 United States          0
12 1991              Frank Dellarosa      21.50 United States          1
13 1992              Frank Dellarosa      19.00 United States          0
14 1993                 Mike DeVito      17.00 United States          0
15 1994                 Mike DeVito      20.00 United States          0
16 1995               Edward Krachie      19.50 United States          0
```

Lab 1: Bar Chart

- 데이터 접근
  - `hotdogs$Winner`

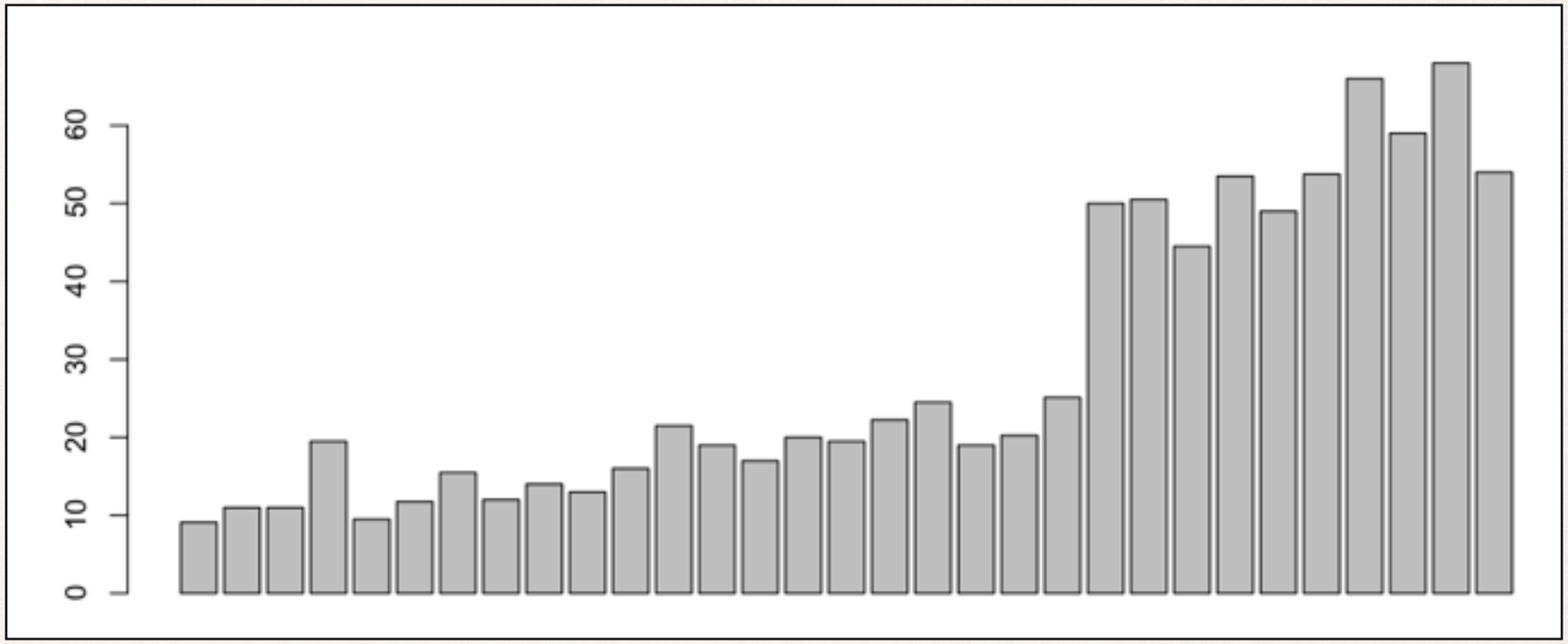
```
> hotdogs$Winner
 [1] Paul Siederman & Joe Baldini Thomas DeBerry
 [3] Steven Abrams                Luis Llamas
 [5] Birgit Felden                Oscar Rodriguez
 [7] Mark Heller                  Don Wolfman
 [9] Jay Green                    Jay Green
[11] Mike DeVito                  Frank Dellarosa
[13] Frank Dellarosa              Mike DeVito
[15] Mike DeVito                  Edward Krachie
[17] Edward Krachie               Hirofumi Nakajima
[19] Hirofumi Nakajima            Steve Keiner
[21] Kazutoyo "The Rabbit" Arai   Takeru Kobayashi
[23] Takeru Kobayashi             Takeru Kobayashi
[25] Takeru Kobayashi             Takeru Kobayashi
[27] Takeru "Tsunami" Kobayashi   Joey Chestnut
[29] Joey Chestnut                Joey Chestnut
[31] Joey Chestnut
```

  - `hotdogs$DogsEaten`

```
> hotdogs$DogsEaten
 [1]  9.10 11.00 11.00 19.50  9.50 11.75 15.50 12.00 14.00 13.00 16.00 21.50
[13] 19.00 17.00 20.00 19.50 22.25 24.50 19.00 20.25 25.13 50.00 50.50 44.50
[25] 53.50 49.00 53.75 66.00 59.00 68.00 54.00
```

Lab 1: Bar Chart

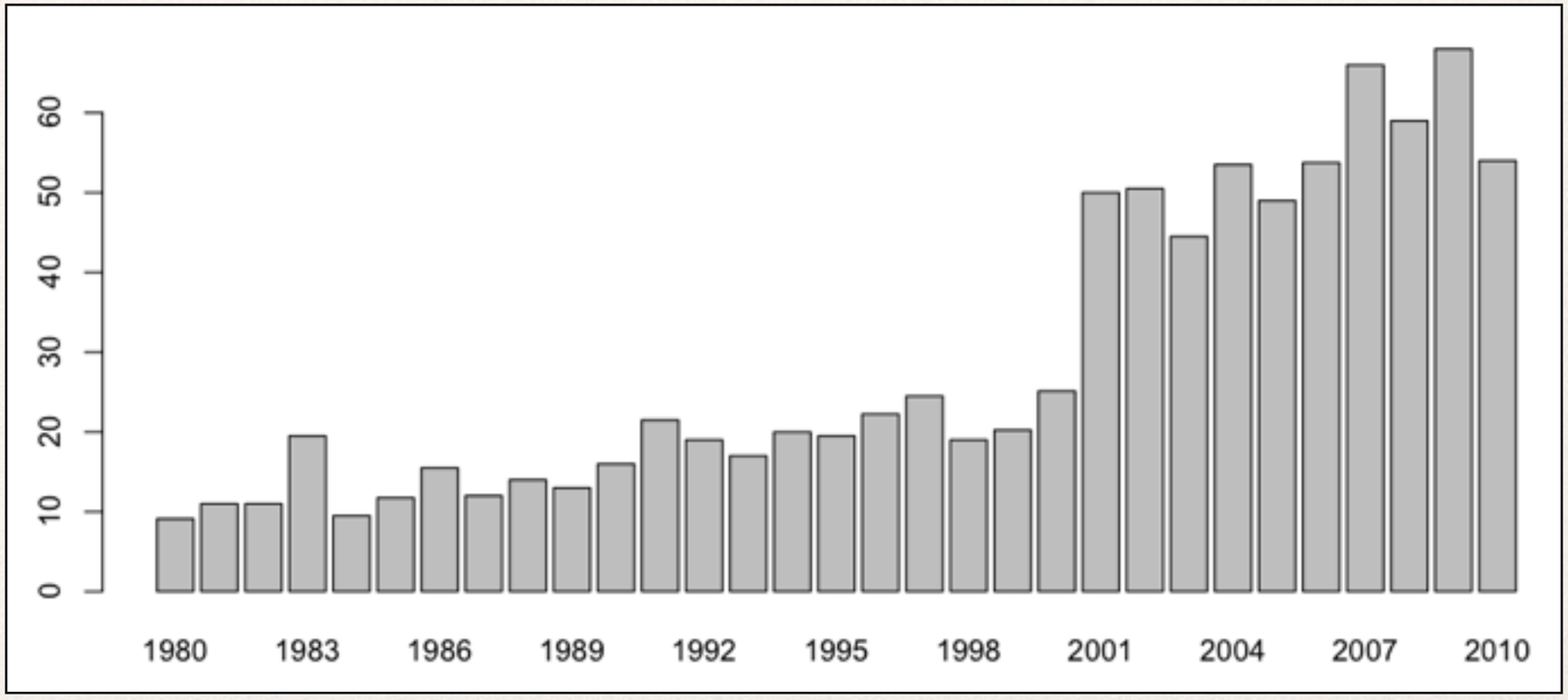
- Bar Chart 그리기
  - `barplot(hotdogs$DogsEaten)`



Lab 1: Bar Chart

- Bar Chart 그리기
  - 연도 레이블 추가

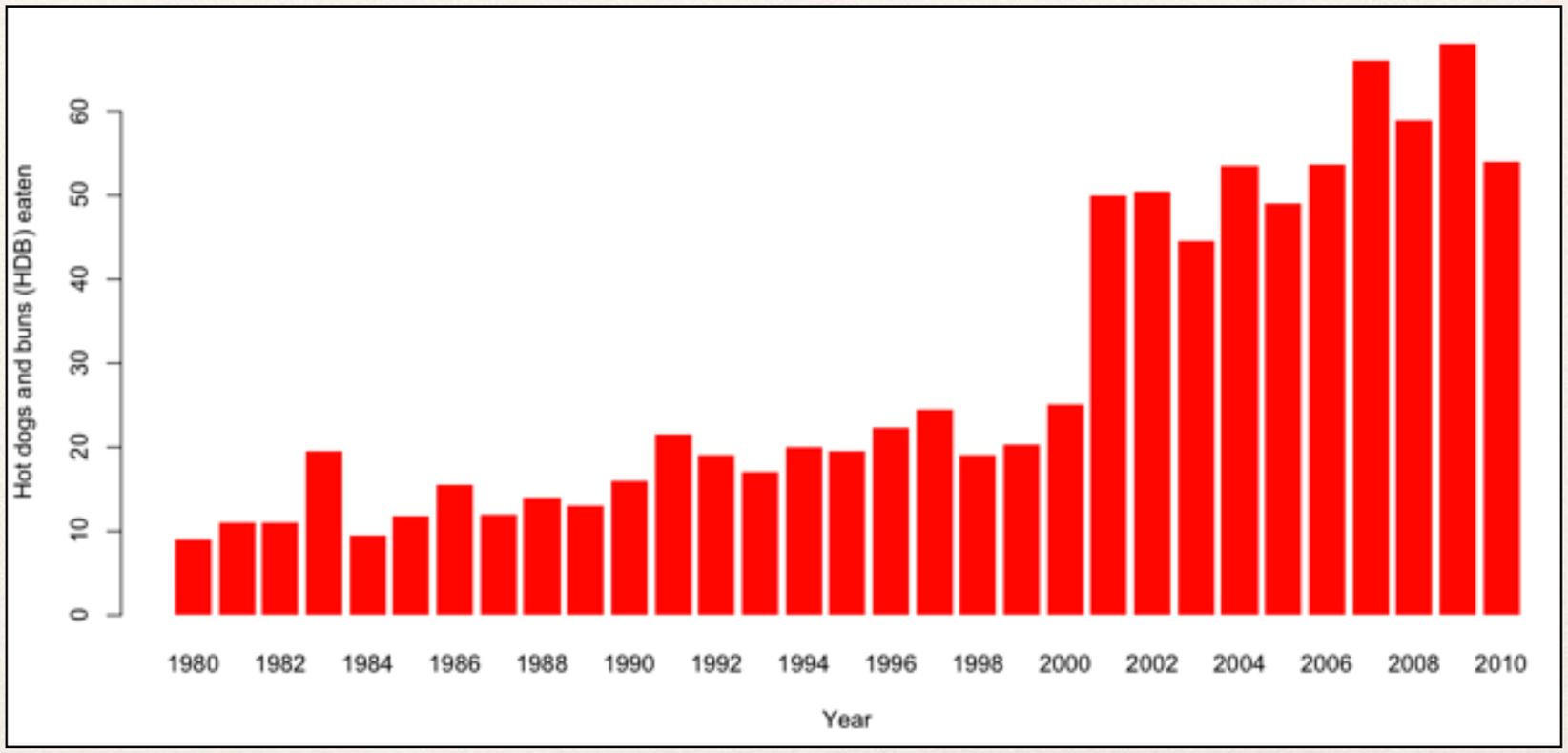
```
barplot(hotdogs$DogsEaten,
names.arg=hotdogs$Year)
```



Lab 1: Bar Chart

- Bar Chart 그리기
  - 바의 색상 변경 및 x, y 레이블 그리기

```
barplot(hotdogs$DogsEaten,
names.arg=hotdogs$Year, col="red",
border=NA, xlab="Year", ylab="Hot dogs
and buns (HDB) eaten")
```



Lab 1: Bar Chart

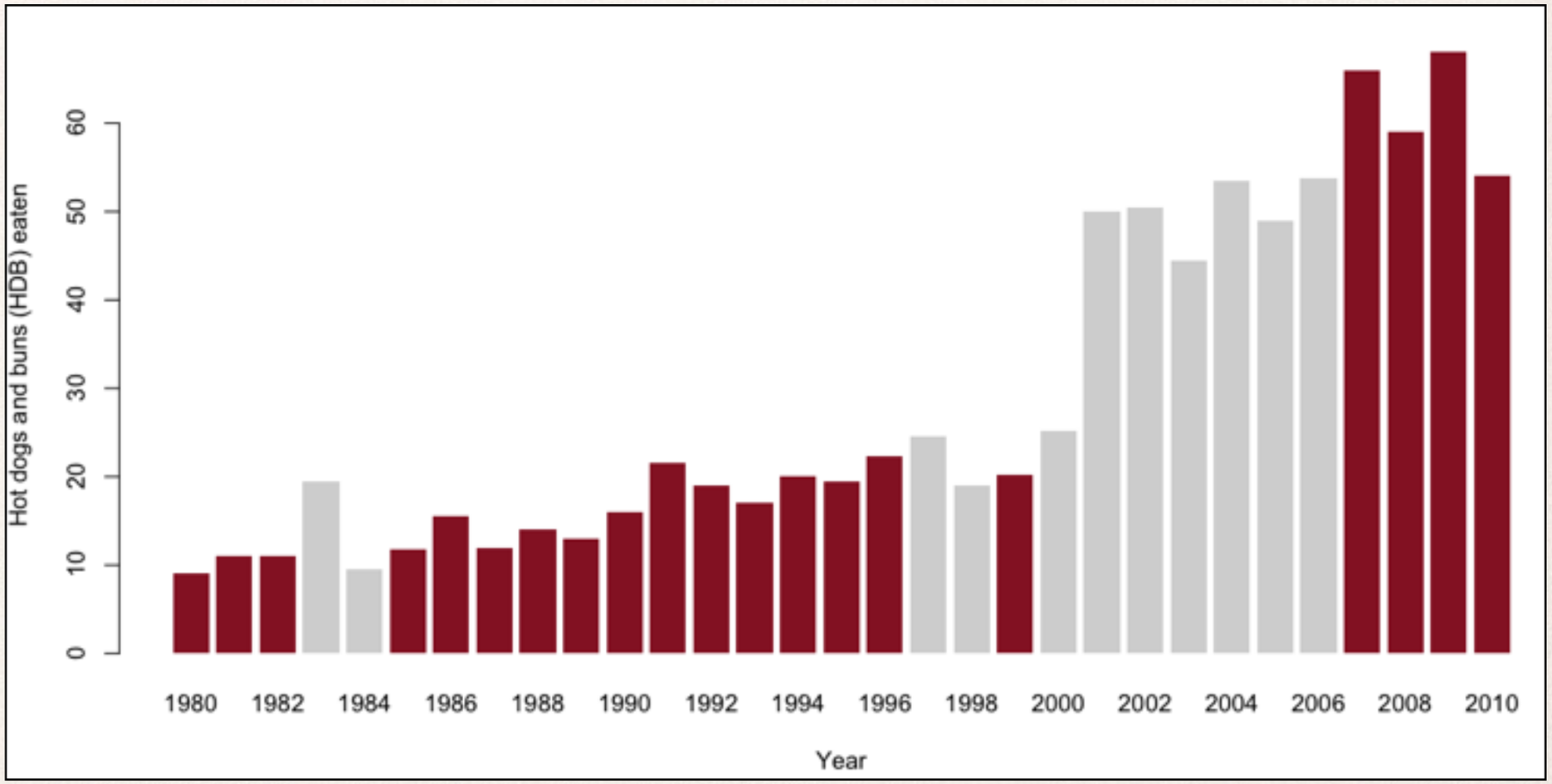
- Bar Chart 그리기
  - color encoding을 통해 미국인과 그 이외를 구분하여 색칠
    - col 옵션에 채워 넣을 값을 벡터로 만들기
    - color value 대신에 벡터 변수 대입

```
fill_colors <- c()
for( i in 1:length(hotdogs$Country) ) {
    if (hotdogs$Country[i] == "United States") {
        fill_colors <- c(fill_colors, "#821122")
    } else {
        fill_colors <- c(fill_colors, "#cccccc")
    }
}

barplot(hotdogs$DogsEaten, names.arg=hotdogs$Year,
col=fill_colors, border=NA, xlab="Year", ylab="Hot
dogs and buns (HDB) eaten")
```

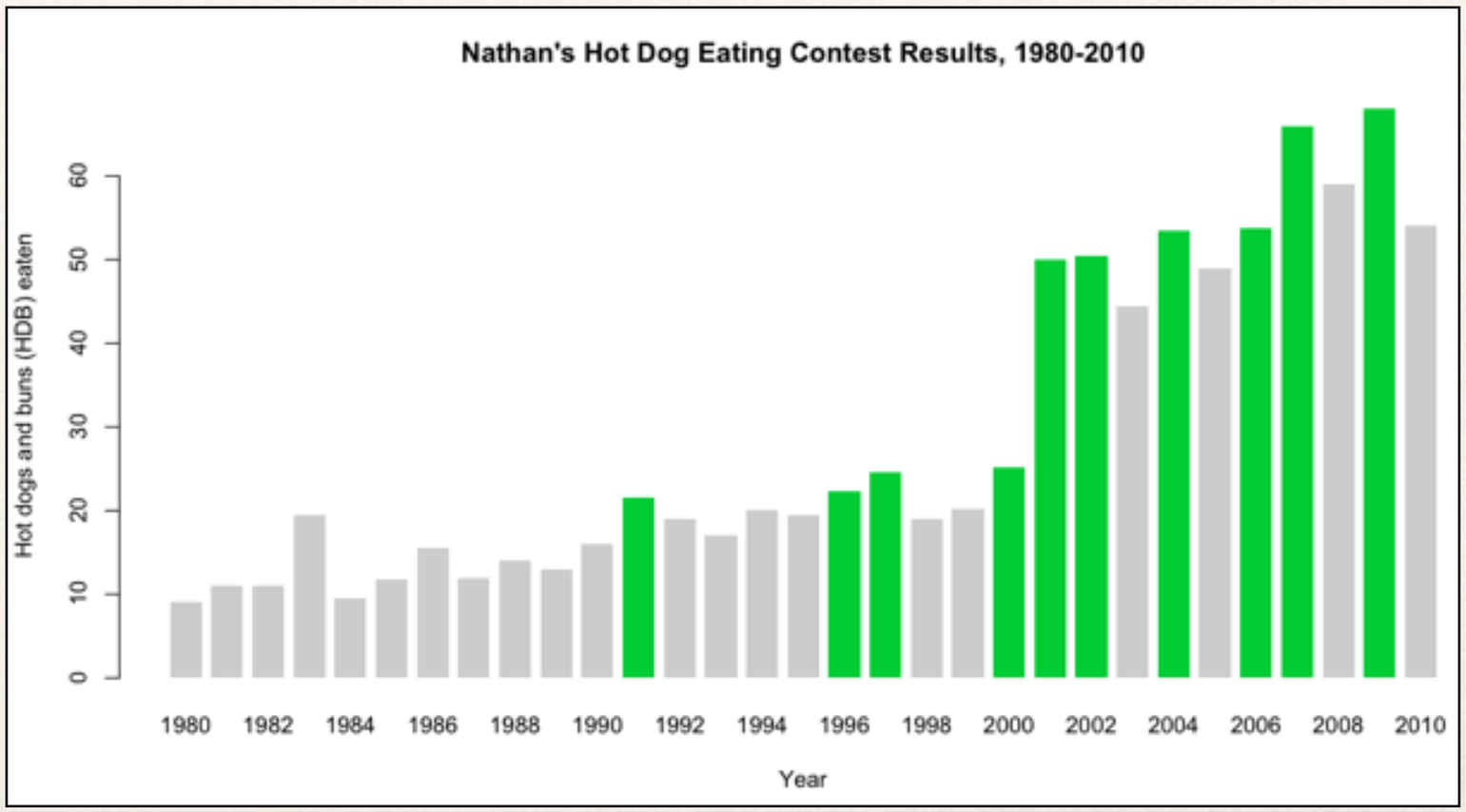
Lab 1: Bar Chart

- Bar Chart 그리기
  - color encoding을 통해 미국인과 그 이외를 구분하여 색칠



Lab 1: Bar Chart

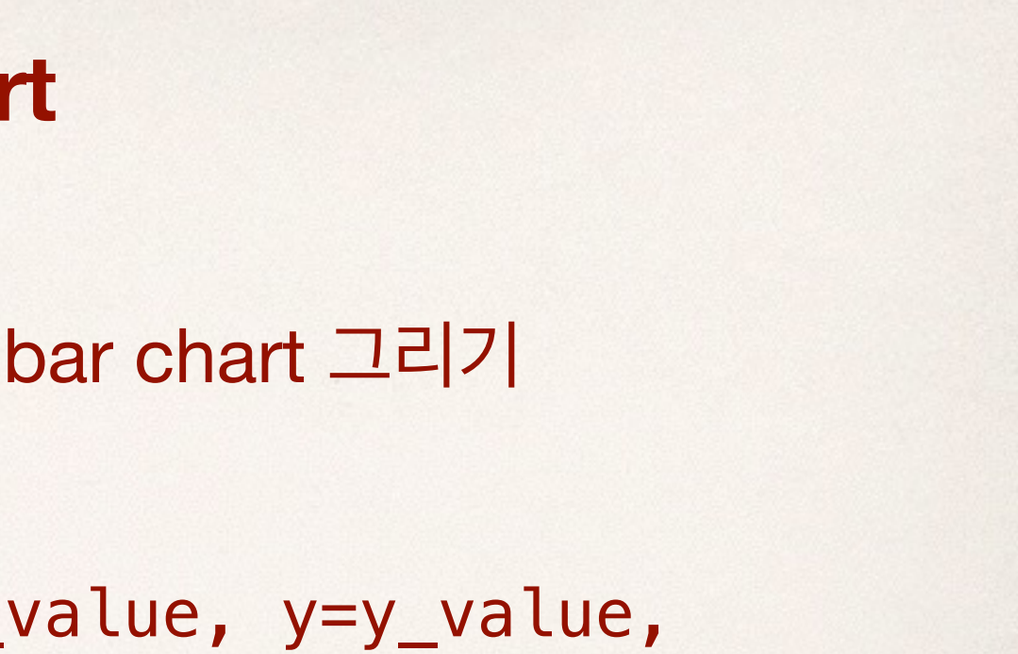
- Bar Chart 그리기
  - 신기록이 나온 해의 bar를 녹색(00cc33)으로 표현
  - bar 사이의 간격을 0.3으로 설정 (hint: space 옵션)
  - Nathan's Hot Dog Eating Contest Results, 1980-2010 이라는 타이틀 표시 (hint: main 옵션)



Lab 2: Stacked Bar Chart

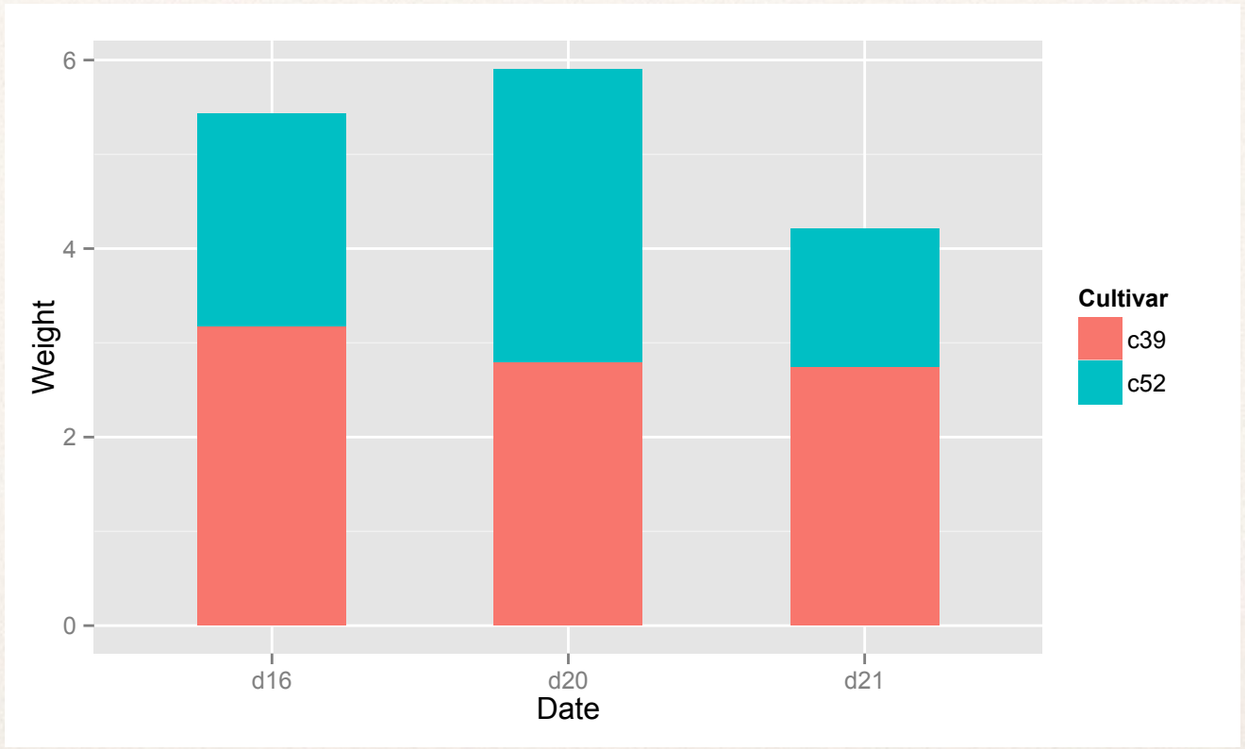
- ggplot2을 사용하여 stacked bar chart 그리기
  - `library(ggplot2)`
  - `ggplot(data, aes(x=x_value, y=y_value, fill=fill_value)`
  - `+ geom_bar (or geom_line, geom_rect, geom_point …)`

- **`ggplot(cabbage_exp, aes(x=Date, y=Weight, fill=Cultivar)) + geom_bar(position="dodge")`**



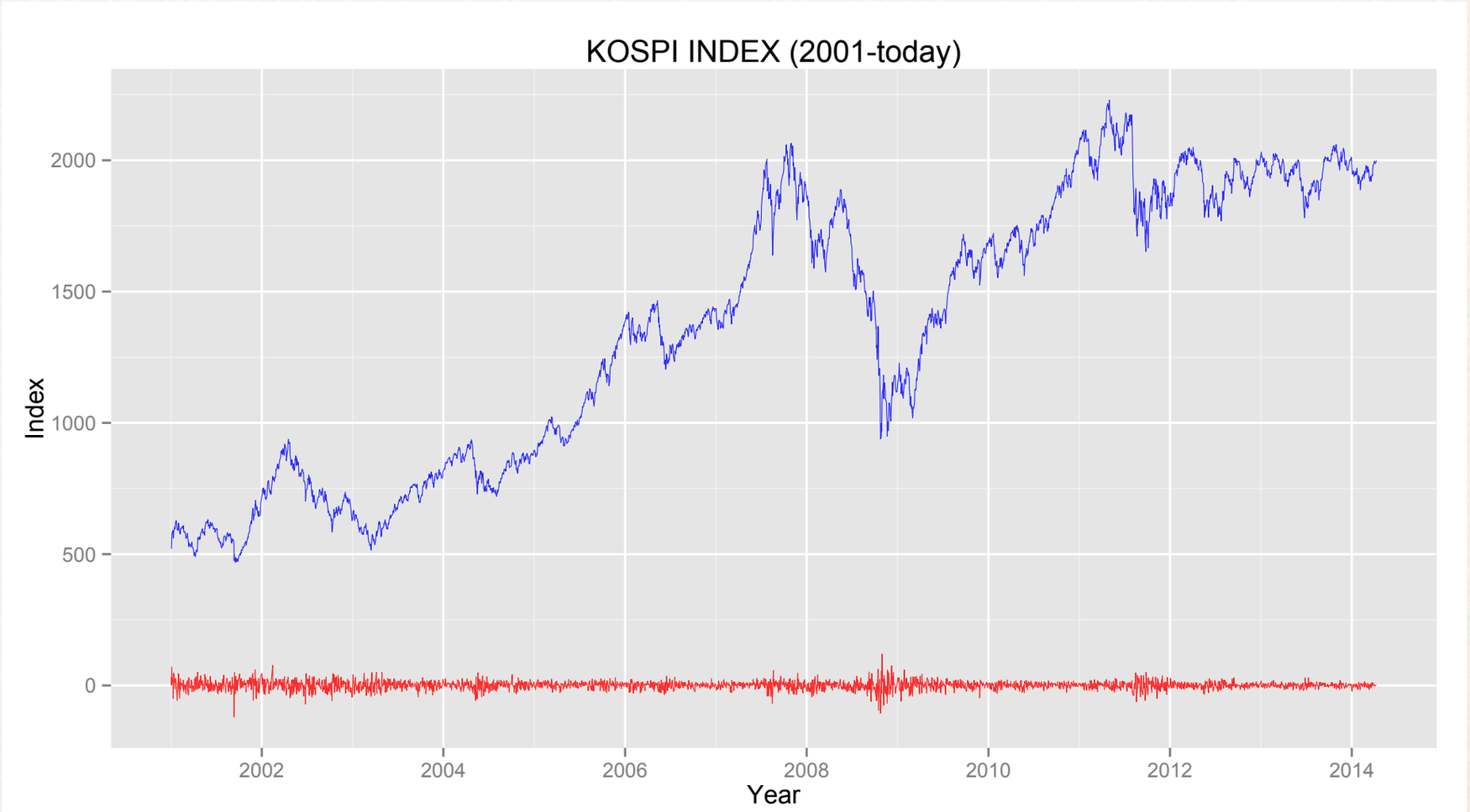
Lab 2: Stacked Bar Chart

- ggplot2을 사용하여 stacked bar chart 그리기
  - **`ggplot(cabbage_exp, aes(x=Date, y=Weight, fill=Cultivar)) + geom_bar(stat="identity")`**
  - add **`width=0.5`** to geom_bar



Lab 3: Line Graph 01

- ggplot2 를 이용한 라인그래프
- 2001년부터 현재까지 KOSPI 주가 데이터를 이용



Lab 3: Line Graph 01

- KOSPI 주가지수 수집, excel 파일로 저장
  - http://www.krx.co.kr/m1/m1_1/m1_1_4/JHPKOR01001_04.jsp
    - 2001.xls ~ 2014.xls
- Load data
  - `library(gdata) #엑셀파일 import`
  - `data_2001 <- read.xls("Data/KOSPI/2001.xls", sheet=1, header=FALSE)`
  - `… data_2014`

Lab 3: Line Graph 01

- Merge data
  - `data <- rbind(data_2001, data_2002, data_2003, data_2004, data_2005, data_2006, data_2007, data_2008, data_2009, data_2010, data_2011, data_2012, data_2013, data_2014)`
- Add header
  - `colnames(data) <- c("date","index","contrast","change","m_value","h_value","l_value","volume_wof","volume_wf","money_wof","money_wf","aggr_value_wof","aggr_value_wf")`

Lab 3: Line Graph 01

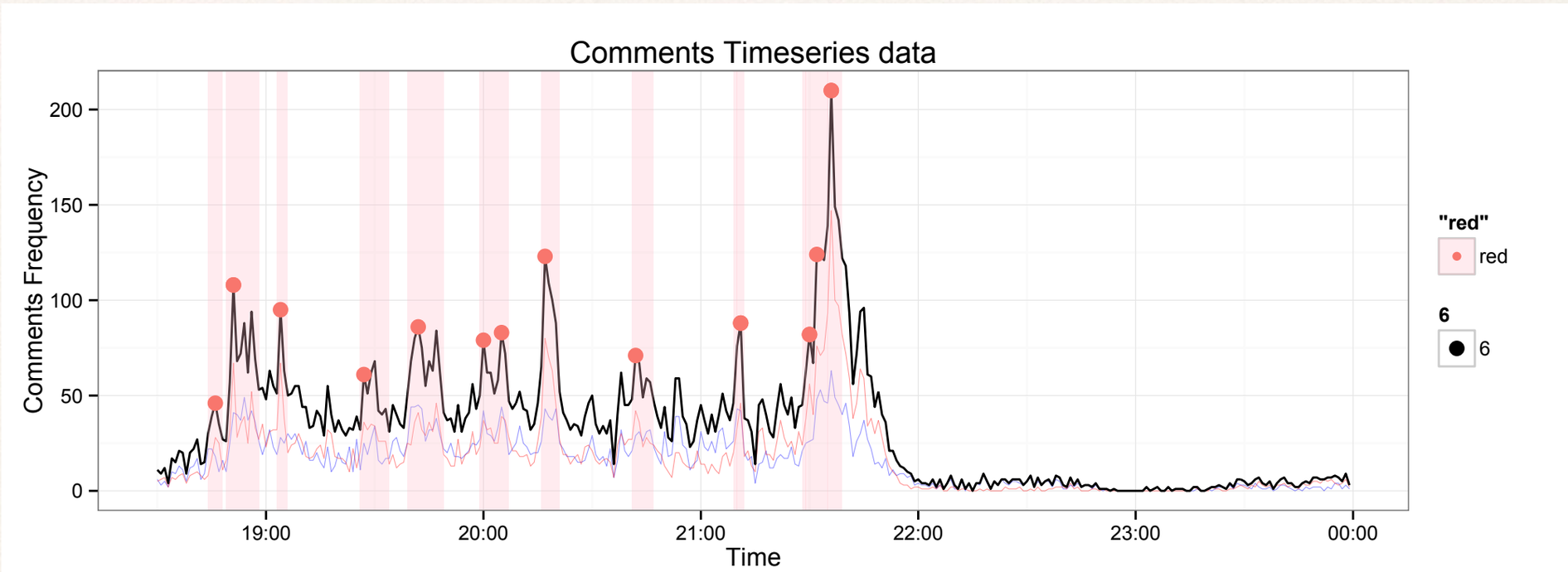
- Data cleaning
  - `data$date <- as.Date(as.character(data$date),format="%Y/%m/%d")`
  - `data$index <- as.numeric(gsub("\\,","",data$index))`

Lab 3: Line Graph 01

- Plot graph
  - ggplot() +

```
geom_line(aes(x=data$date, y=data$index),
color="blue", size=0.2, alpha=0.8) +
geom_line(aes(x=data$date, y=data
$change*10), color="red", size=0.2,
alpha=0.8) +
labs(title = "KOSPI INDEX (2001-today)",
x = "Year", y = "Index")
```

Lab 4: Line Graph 02



- Load data
  - `peaksfinding_result <- read.csv("Data/baseball-sns-peaks.csv",header=T)`
  - `timeseries_data <- read.csv(“Data/baseball-sns.csv",header=T,stringsAsFactor=F)`

Lab 4: Line Graph 02

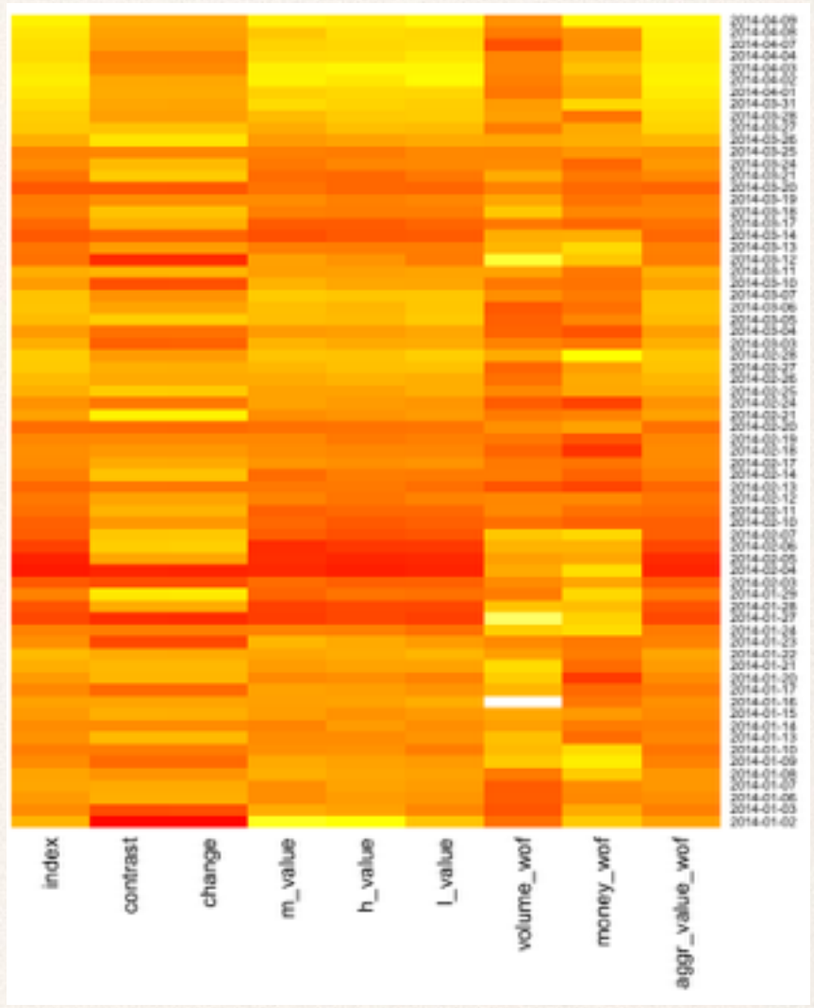
- Data type conversion
  - `timeseries_data$time <- as.POSIXct(timeseries_data$time,format = "%m/%d/%y %H:%M")`

Lab 4: Line Graph 02

- Plot graph
  - ```
    ggplot() +
    geom_line(aes(x=timeseries_data$time, y=timeseries_data
    $away), color="blue", size=0.2, alpha=0.4) +
    geom_line(aes(x=timeseries_data$time, y=timeseries_data
    $home), color="red", size=0.2, alpha=0.4) +
    geom_line(aes(x=timeseries_data$time, y=timeseries_data
    $total)) + theme_bw() +
    labs(title = "Comments Timeseries data", x = "Time", y =
    "Comments Frequency") +
    geom_rect(data=peaksfinding_result,aes(xmin=as.POSIXct(star
    t_time,origin="1970-01-01"),xmax=as.POSIXct(end_time,origin
    ="1970-01-01"),ymin=-Inf, ymax=+Inf,
    color="red"),fill='pink', alpha=0.3, linetype="blank") +
    geom_point(data=peaksfinding_result,aes(x=as.POSIXct(peak_t
    ime,origin="1970-01-01"),y=peak_fq, color="red",size=6))
    ```
    ```

Lab 5: Heat Map

- Load data
  - KOSPI 데이터가 너무 양이 많으므로 2014년 데이터만 사용
- 2014.xls 만 로딩하거나,
  전체 로딩된 데이터에서
  필요한 부분의 subset을 추출



- `data.subset <- subset(data, as.Date(date) >= '2014-01-01' & as.Date(date) <= '2014-12-31')`

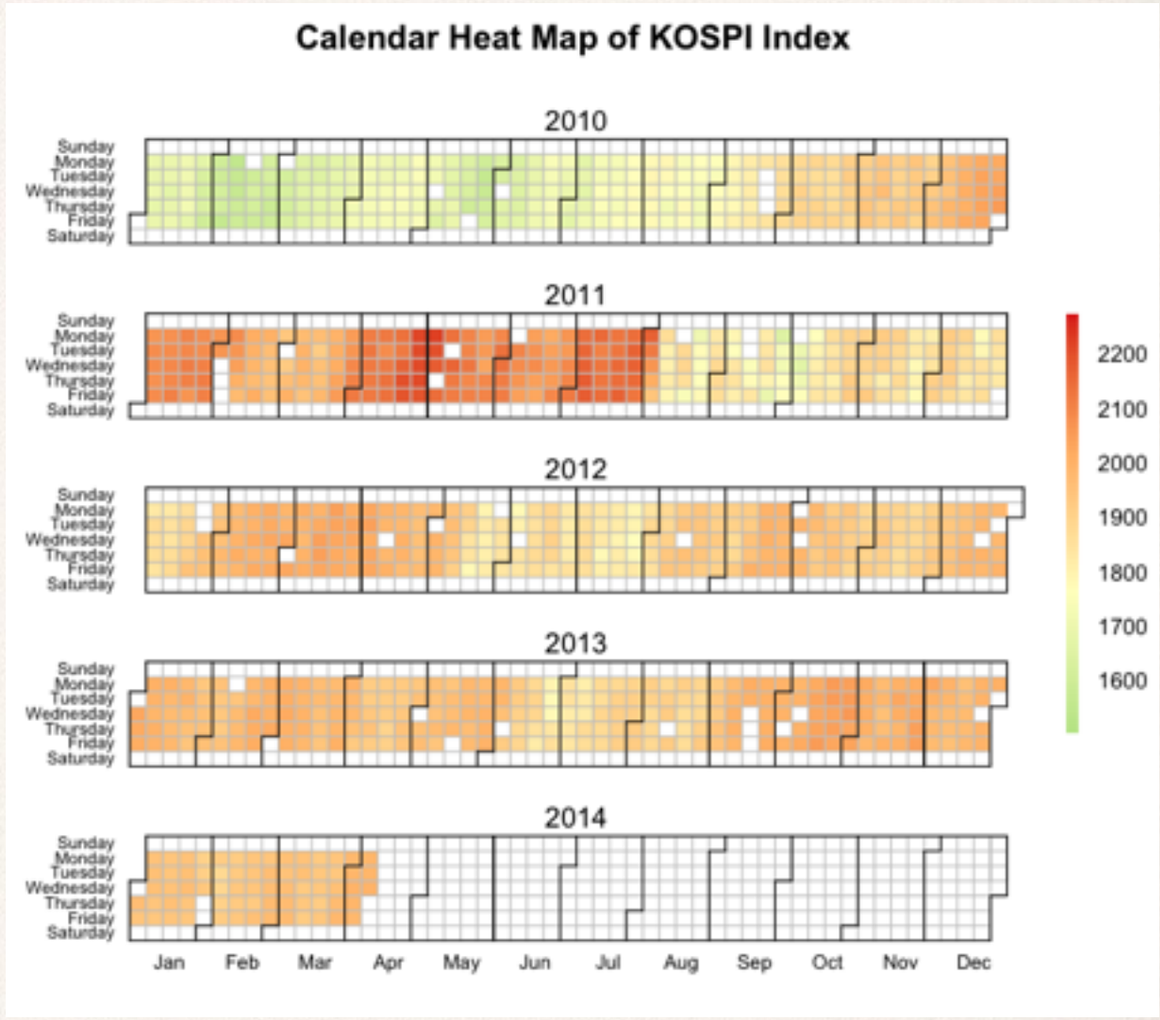
Lab 5: Heat Map

- Row names
  - Heatmap 의 y축을 나타낼 row names 설정
  - `row.names(data.subset) <- data.subset$date`
- Filtering columns
  - 모든 컬럼이 필요하지 않기 때문에 필요한 부분만 선택하여 사용
  - `data.subset.filtered <- subset(data.subset, select=c(index, contrast, change, m_value, h_value, l_value, volume_wof, money_wof, aggr_value_wof))`

Lab 5: Heat Map

- Data frame 을 Data matrix 로 변환
  - Data frame: 다양한 타입의 변수 저장 가능
  - Data matrix: 한 종류의 변수만 저장 가능
    - `data.matrix <- data.matrix(data.subset.filtered)`
- Heatmap drawing
  - `data.heatmap <- heatmap(data.matrix, Rowv=NA, Colv=NA, col=heat.colors(256), scale="column", margins=c(5,10))`

Lab 6: Calendar Heat Map



- Calendar Heat Map
  - 날짜별로 일목요연하게 정리되어 있는 heat map
  - 그리기가 까다로움

    → Paul Bleicher 가 작성한 calendarHeat 함수를 사용
    - http://blog.revolutionanalytics.com/2009/11/charting-time-series-as-calendar-heat-maps-in-r.html

Lab 6: Calendar Heat Map

- Subset data
  - `data.subset <- subset(data, as.Date(date) >= '2010-01-01' & as.Date(date) <= '2014-12-31')`
- Draw a heat map
  - `calendarHeat(dates=data.subset$date, values=data.subset$index, color="g2r", varname="KOSPI Index")`

Questions...?