2 - Advanced Graphics in R

04 - Dates, Times, and Groups

Iowa State University

Outline

- Dates and Times
- `lubridate` package

Dates and Times

- Dates and times are deceptively tricky to work with
- Formats - 02/05/2012 is February 5 or May 2?
- Time Zones
- POSIXct and POSIXlt format in R is difficult to work with

`lubridate` package

- available from CRAN (`install.packages("lubridate")`)
- Written by Garett Grolemund and Hadley Wickham
- associated paper
  JSS: Dates and Times made easy with lubridate
  `http://www.jstatsoft.org/v40/i03/paper`

Instants of time

- one moment in time, usually named, e.g.

```
now()
## [1] "2014-10-01 15:54:36 CDT"
```

- lubridate turns strings into instants with functions that have y, m, and d in their names

```
ymd("2013-05-14")
## [1] "2013-05-14 UTC"

mdy("05/14/2013")
## [1] "2013-05-14 UTC"

dmy("14052013")
## [1] "2013-05-14 UTC"

ymd_hms("2013:05:14 14:50:30")
## [1] "2013-05-14 14:50:30 UTC"
```

- Order matters!

Your Turn

- The data set chicago.csv contains records for every flight departing from Chicago O'Hare in June 2008
- Parse the Date variable into a Date-Time Object

Working with instants

- Standard arithmetic operations now work on dates:

```
now() > ymd("1970-01-01")
## [1] TRUE
```

```
now() - ymd("1970-01-01")
## Time difference of 16345 days
```

- functions for extracting pieces of dates:

```
month(now())
## [1] 10
```

```
wday(now())
## [1] 4
```

```
wday(now(), label=TRUE)
## [1] Wed
## 7 Levels: Sun < Mon < Tues < Wed < ... < Sat
```

- What's your age in days?

Accessor functions

Component	Function
Year	`year()`
Month	`month()`
Day of the year	`yday()`
Day of the month	`mday()`
Day of the week	`wday()`
Hour	`hour()`
Minute	`minute()`
Second	`second()`
Time zone	`tz()`

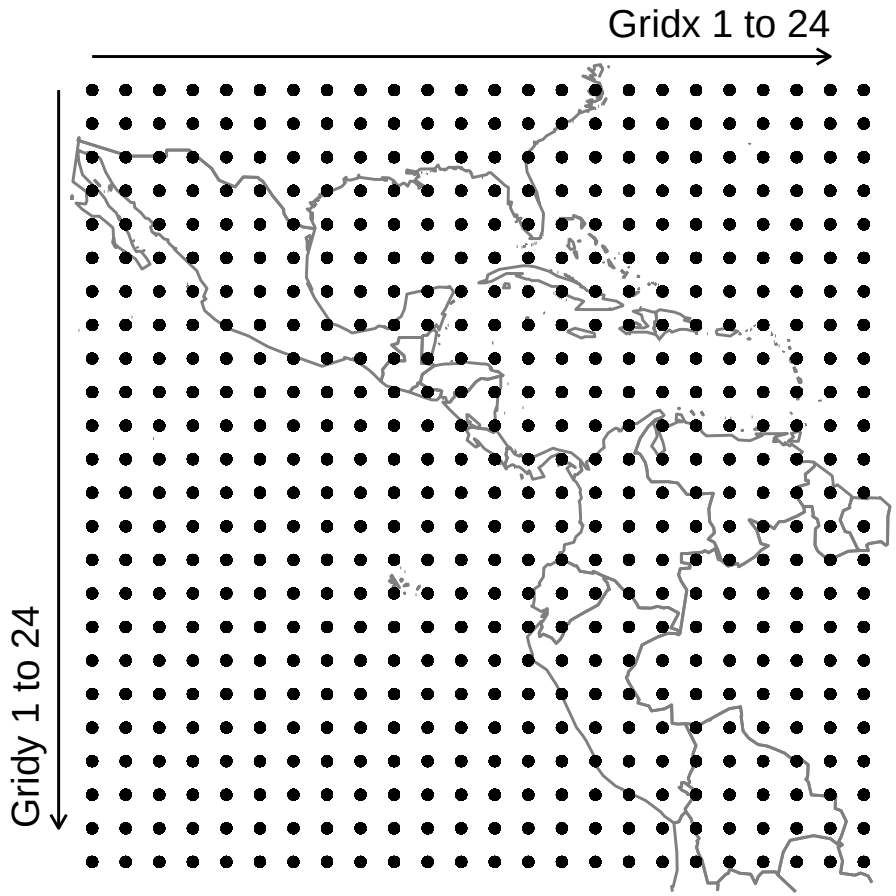
What day of the year were you born?

Your Turn

- For the chicago.csv data set find out whether day of the week has an impact on departure delays (`DepDelay`) (FAA defines a delay as 15 minutes or more) You could draw a boxplot by day of the week, or sum delays by day of the week, ...
- How many Sundays or Mondays did June 2008 have? Give a breakdown of week day frequencies. Does that change your initial answer?

NASA Meterological Data

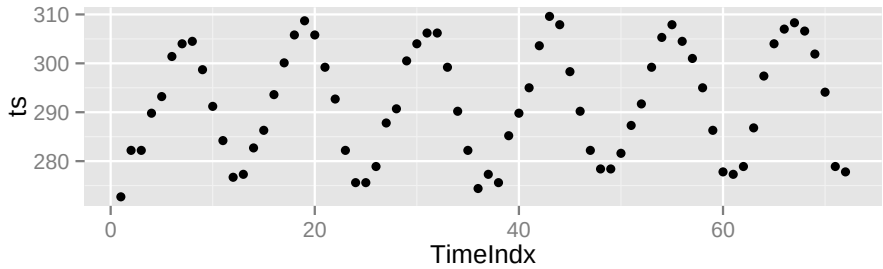
- `ggplot2` can work nicely with time objects provided by lubridate
- 24 x 24 grid across Central America
- Satellite captured data: temperature (ts), near surface temperature (`tsa`), pressure (`ps`), ozone (o3), cloud coverage: low (`ca_low`), medium (`ca_med`), high(`ca_high`)
- for each location monthly averages for Jan 1995 to Dec 2000



What is a Time Series?

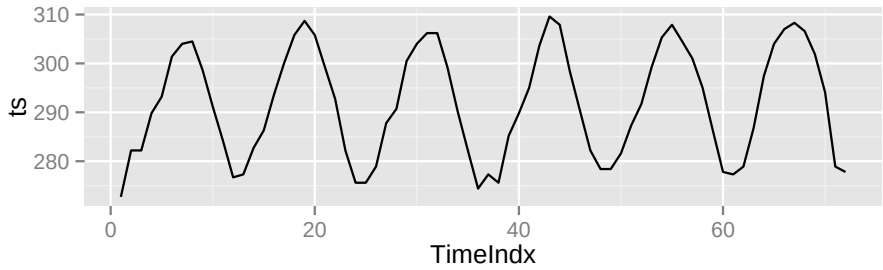
- For each location multiple measurements

```
qplot(TimeIndx, ts, geom="point",
      data=subset(nasa, (Gridx==1)&
                        (Gridy==1)))
```



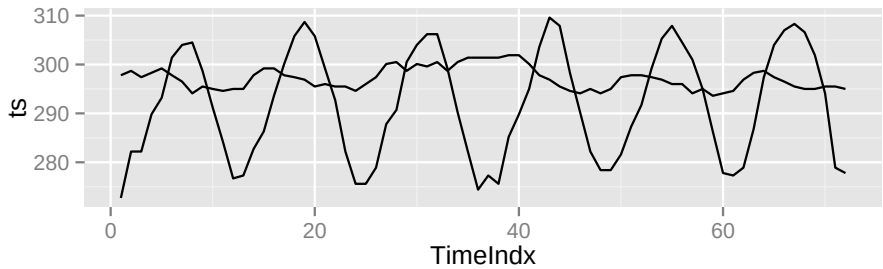
- connected by a line

```
qplot(TimeIndx, ts, geom="line",
      data=subset(nasa, (Gridx==1)&
                        (Gridy==1)))
```



- but only connect the right points

```
qplot(TimeIndx, ts, geom="line",
      data=subset(nasa, (Gridx==1)&
                        (Gridy%in%c(1,15))),
      group=Gridy)
```



Your Turn

- Get a subset of locations, plot a time series for pressure for each location.

  What is the general pattern?

- For all locations, draw individual time series for pressure.

  What do you expect? Are there surprising values? Which are they?

- Introduce a Date (Year + Month) to the nasa data and change the pressure time series plot accordingly.