Primer on time series

Joshua Loftus

July 17, 2015

Outline

- Motivating examples
- A spoonful of theory
- Further reading

ts(): Creating a *time series* object

Google trends: search popularity of "game of thrones"

Read the data and subset to the right part (the .csv file from Google trends is a bit messy)

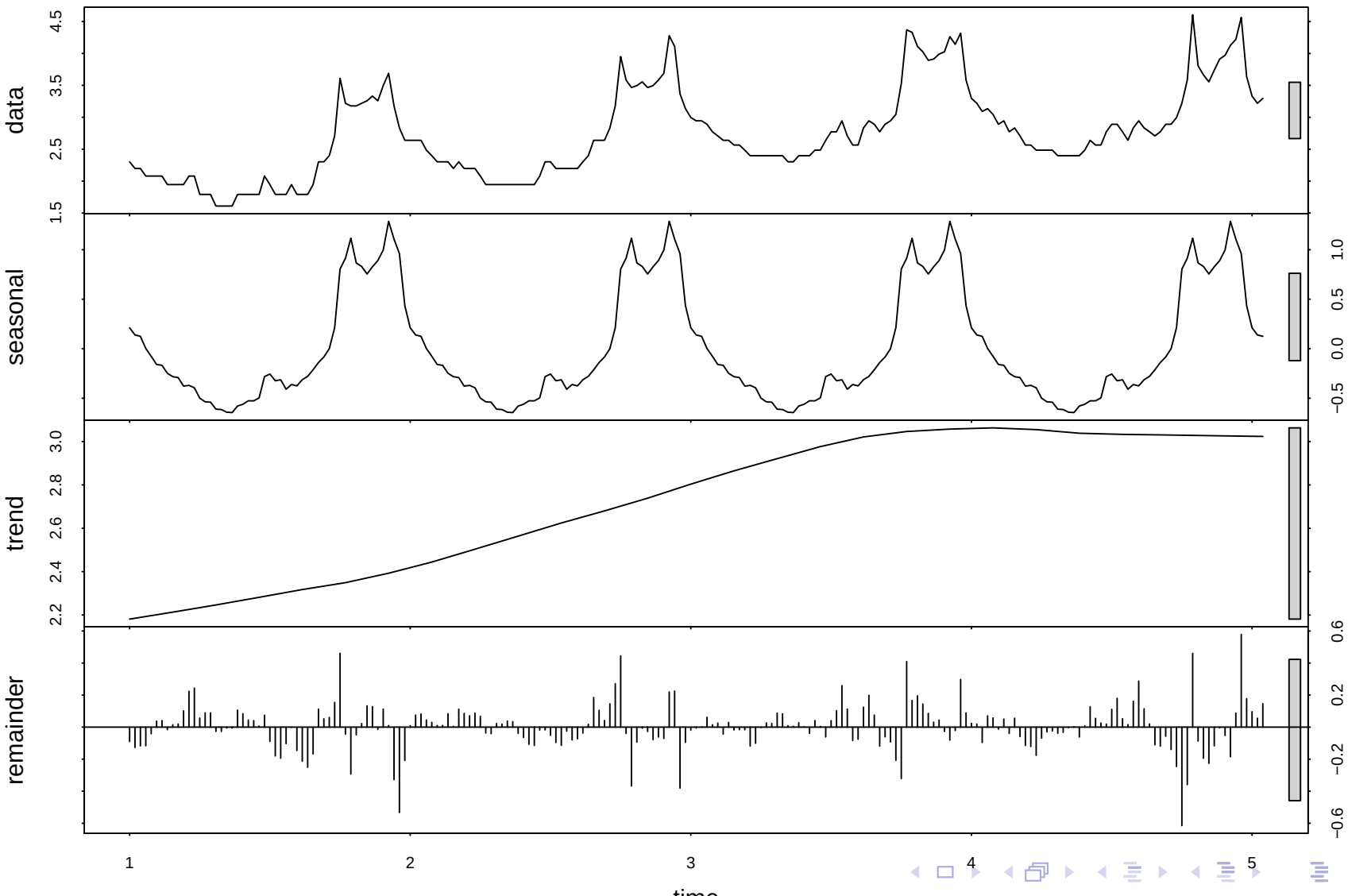
```
setwd("~/Dropbox/work/teaching/consulting/timeseries")
data <- read.csv("GoT.csv", skip = 4, stringsAsFactors = F)
data <- data[1:211,]
data[,2] <- as.numeric(data[,2])
```

The data is given by week. Seasons happen once per year.

```
d <- ts(data[,2], frequency = 52)
```

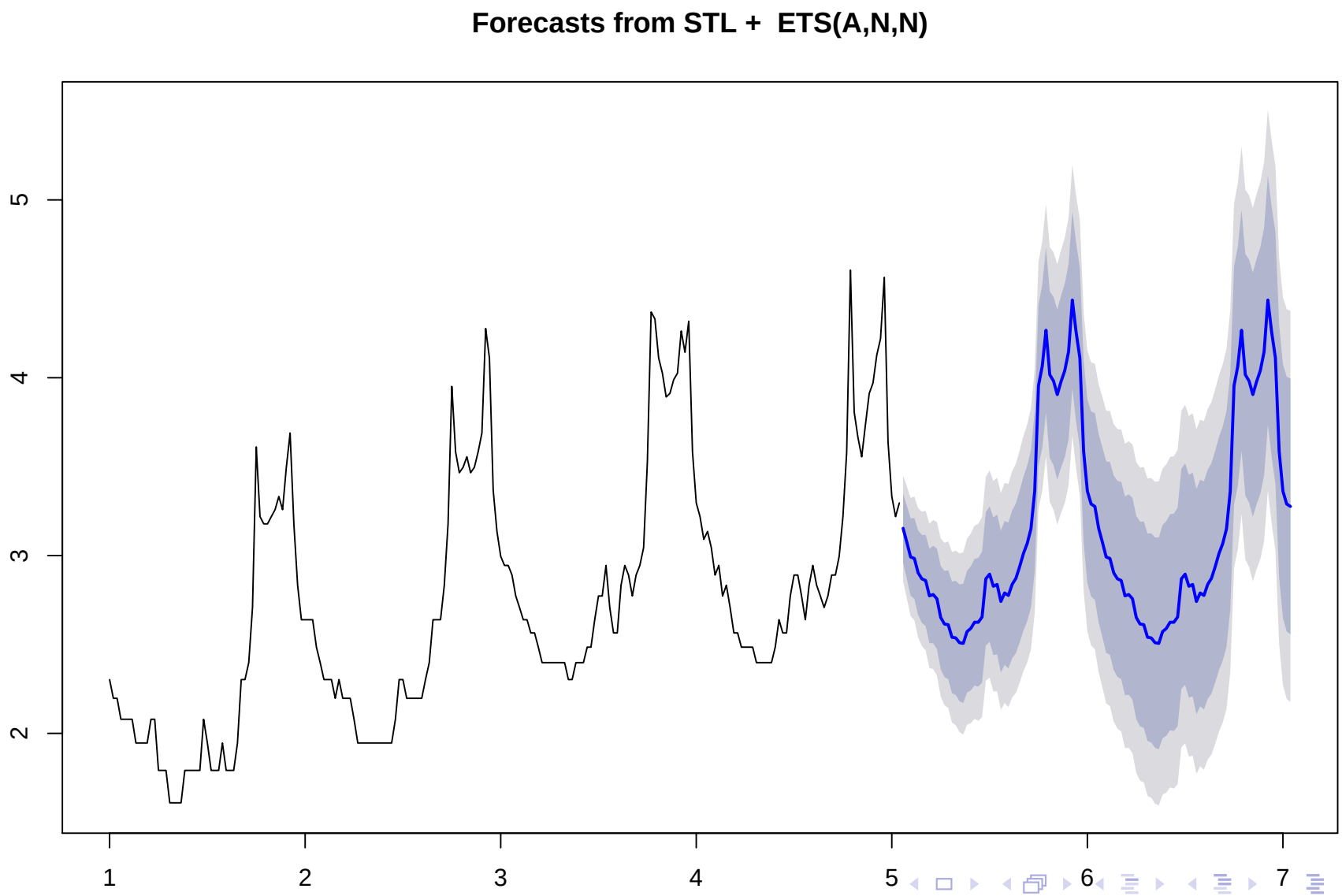
stl(): Seasonal decomposition by Loess

```
fit <- stl(log(d), s.window = "period")
plot(fit)
```



library(forecast): Predicting the future

```
plot(forecast(fit))
```

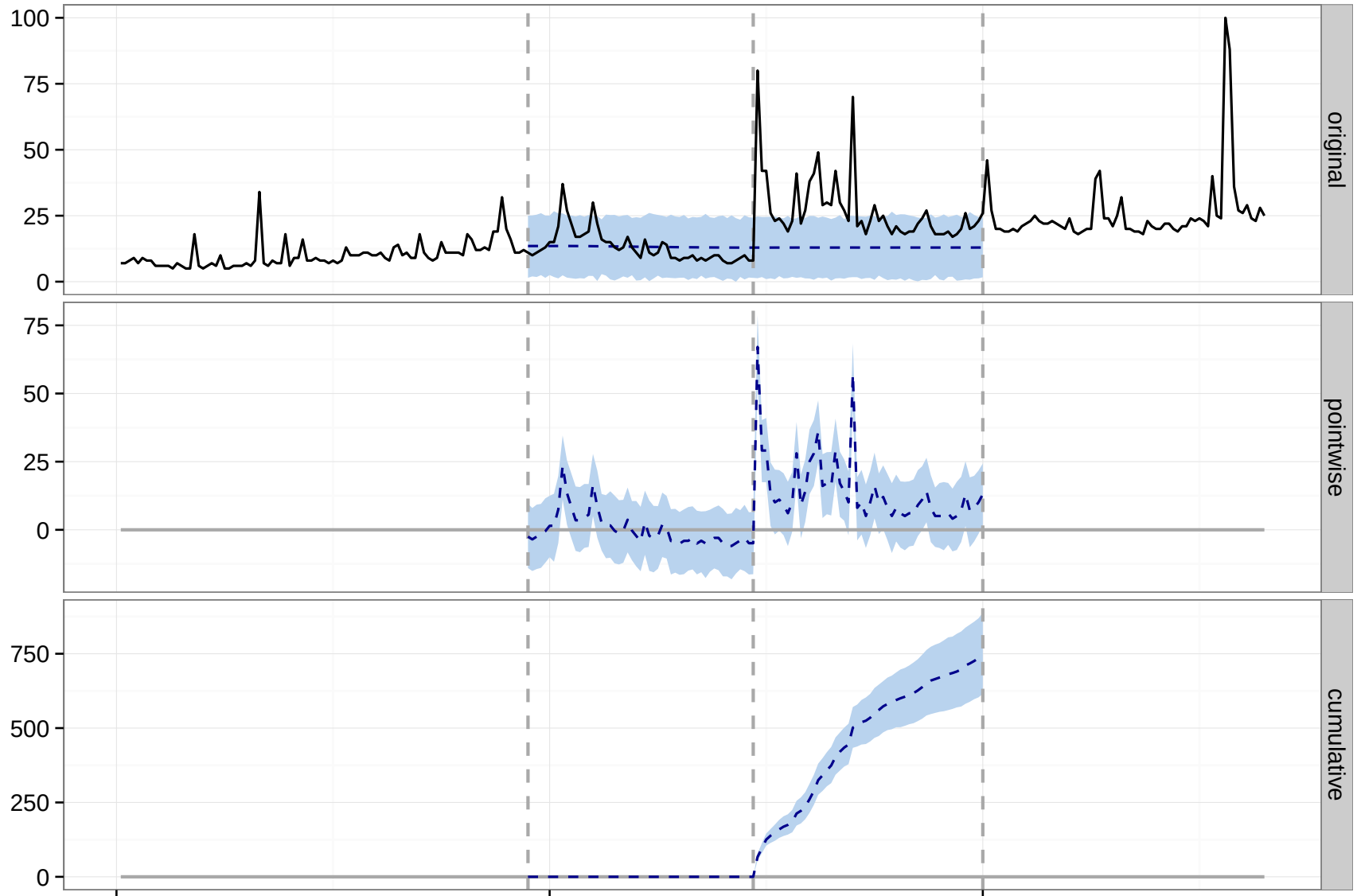


Discontinuity and “causal” inference

- Time series observed before and after an intervention
- If behavior changes dramatically, maybe it was because of the intervention
- Important to rule out other things happening at that time
- Example next slide: search popularity of “Star Wars” before and after Disney purchase announced

library(CausalImpact) developed at Google

```
impact <- CausalImpact(as.numeric(data[,2]), pre.period, po
plot(impact)
```



Stochastic processes

- $\{X_t\}_{t \geq t_0}$
- Collection of random variables indexed by time t, in practice discrete
- Most methods require *stationarity* : $(X_{t_1}, \ldots, X_{t_k})$ has same distribution as $(X_{t_1+h}, \ldots, X_{t_k+h})$
- Transform data by taking logs, differences, to get stationarity
- Many classes of models...

Moving averages and autoregression

- MA(q) moving average: $X_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}$
- Random shock affects future values of X directly
- AR(p) autoregression: $X_t = c + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \epsilon_t$
- Random shock affects future values of X only through past values of X
- AMRA(p,q) autoregessive moving-average
- ARIMA...

Error terms

- ARCH conditional heteroskedasticity: variance of present error depends on observed past errors
- GARCH generalized: also depends on variance of past errors
- e.g. ARIMA/GARCH together quite general (5 parameters)

Further introductory reading

Very short reference
`http://www.statmethods.net/advstats/timeseries.html`

Short, easy tutorial, start in chapter 2
`http://www.statoek.wiso.uni-goettingen.de/veranstaltungen/zeitreihen/sommer03/ts_r_intro.pdf`

Another similar tutorial (I prefer the one above)
`https://a-little-book-of-r-for-time-series.readthedocs.org/en/latest/src/timeseries.html`

A Bayesian approach like "interrupted time series" (developed at Google) `http://www.r-bloggers.com/causalimpact-a-new-open-source-package-for-estimating-caus`

Hidden Markov models (application in genetics)
`http://a-little-book-of-r-for-bioinformatics.readthedocs.org/en/latest/src/chapter10.html`