STAT 485/685
Overview

Richard Lockhart

Simon Fraser University

STAT 485/685 — Fall 2017

Purposes of These Notes

- Outline the contents of this course
- Establish the grading structure
- Outline expectations
- Discuss coverage in text

Me

- Richard Lockhart.
- `lockhart@sfu.ca`.
- `http://www.stat.sfu.ca/~lockhart`.
- K10561.
- Office hours: Monday 4:30 to 5:20; Wednesday 11:30 to 12:30; Thursday 3:30 to 4:20.
- You are free to drop in whenever I am there with the door open.

The Text

- **Text**: by Jonathan Cryer and Kung-sik Chan.
  Time Series Analysis With Applications in R by Jonathan Cryer and Kung-sik Chan.
- **Coverage**: Chapters 1 through 9.

Assignments

- There will be about 8.
- The first is due Wednesday, 20 September by NOON in STAT 485/685 box.
- They have computing components, writing components and the usual mathematical components.
- A string of equations with no words does not constitute an adequate answer.
- You need words and explanation of what you are doing.

Grading

- The assignments are worth 30% of your final grade.
- You are free to talk to each other about them but if I think you are copying each other I will be livid.
- The midterm is worth 30% and there is a final exam worth 40%.
- The midterm is October 19.
- I will be permitting note sheets in exams; amounts to be determined.

Other

- October 9 is Thanksgiving Day: no classes.
- November 13 is a statutory holiday (Remembrance Day is Saturday the 11th): no classes.
- **Web**: I try to put lecture notes on the web in advance.
- I recommend you bring them along to class to add to if you are the sort who likes to take notes.
- I will not be using the lecture notes in class, however. I will write on transparencies in spite of my appalling handwriting.
- You should also feel free to look at other courses I have offered

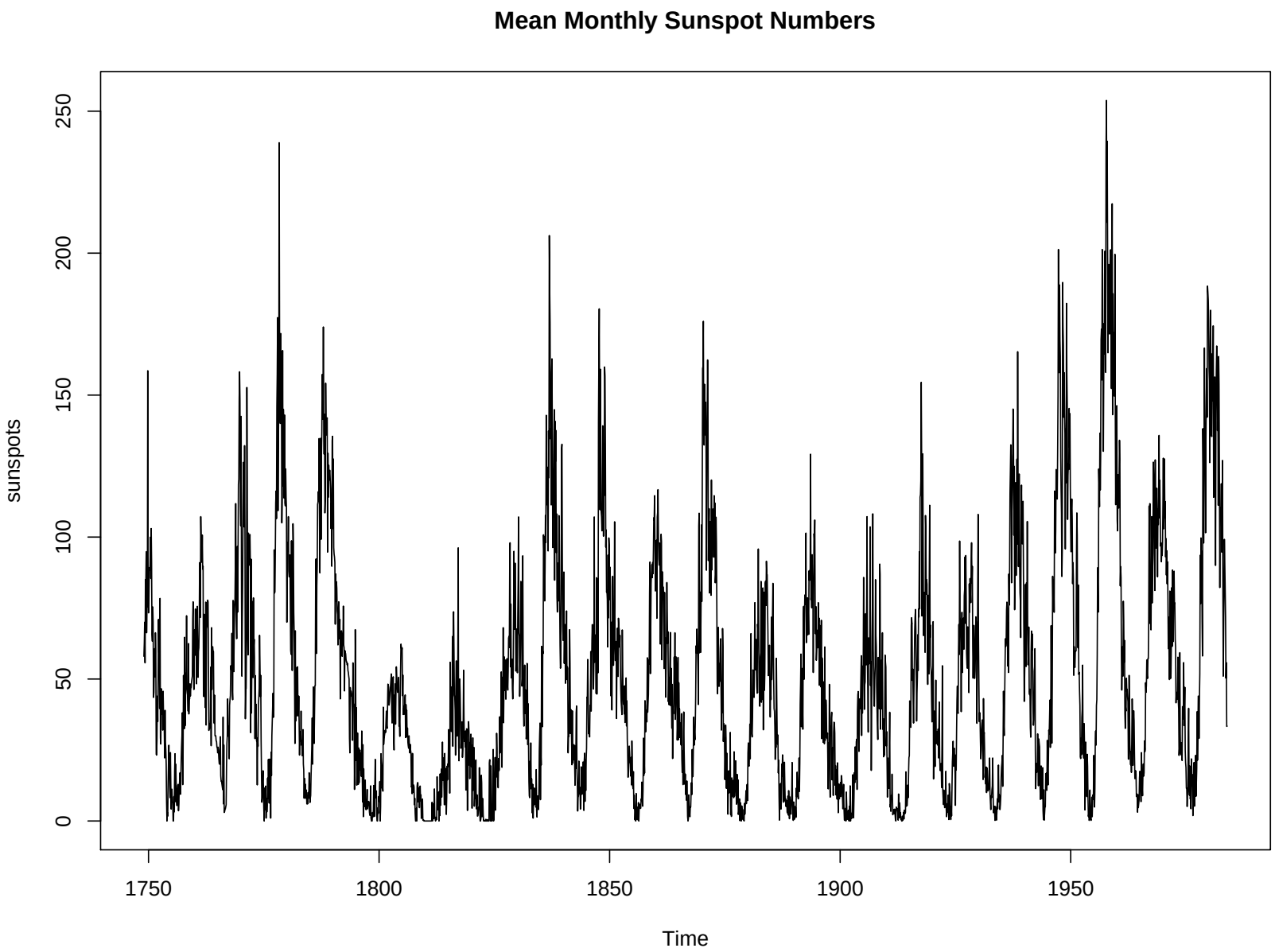
Detailed Course Outline

- This outline will grow as the term goes on, I hope. This is a new course for me though you might occasionally find my STAT 804 notes accessible.

Some Time Series Plots

- I will show some plots then introduce mathematical notation.
- '*Those who forget history are doomed to repeat it.*'
- Can we use the past to help guess the future?
- Random behaviour vs deterministic behaviour.
- Internal drivers vs external drivers.

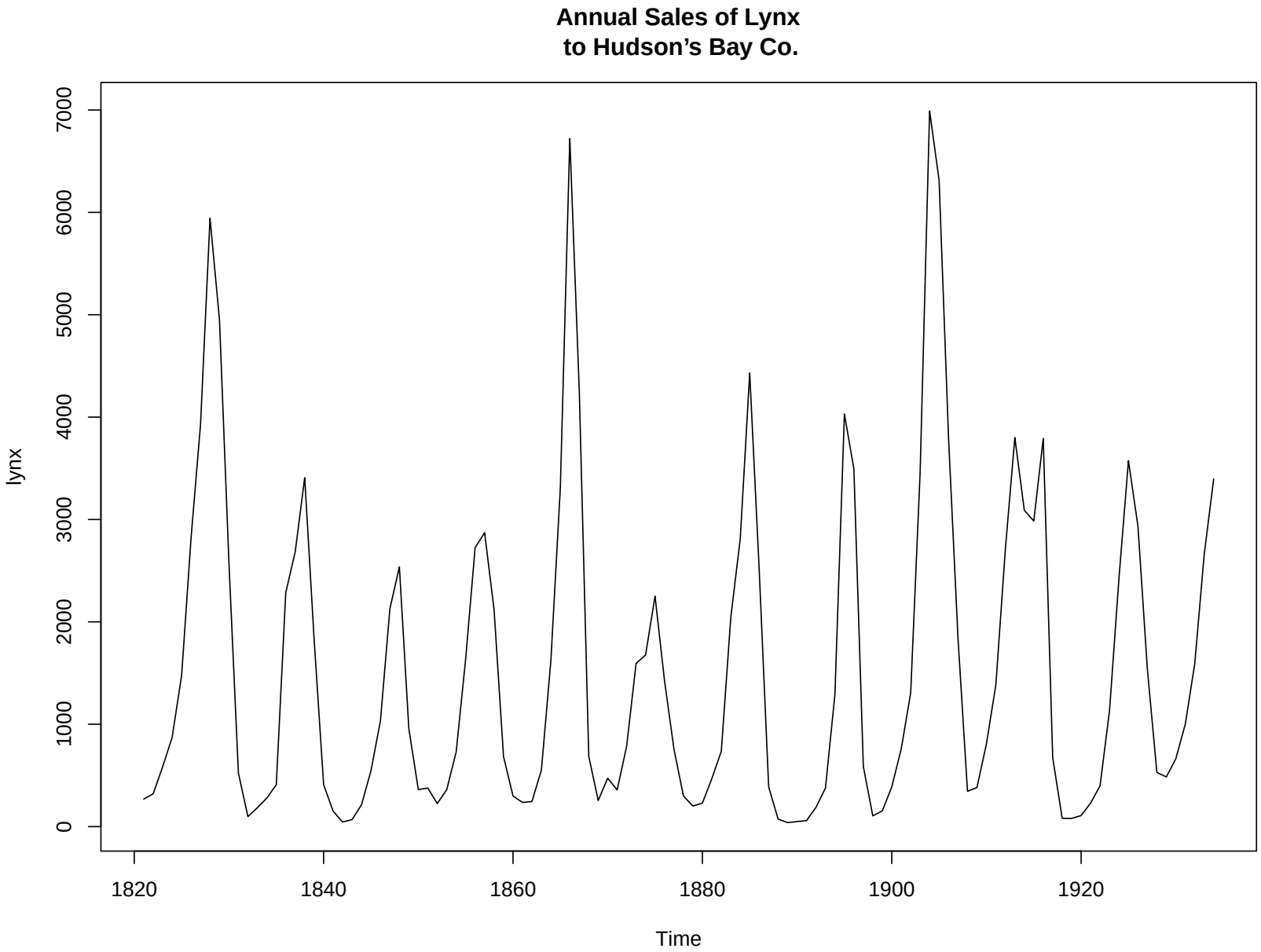
Sunspots


Mean Monthly Sunspot Numbers
sunspots
0
50
100
150
200
250
1750
1800
1850
1900
1950
Time

Comments on Sunspots

- Sunspot data. Each month average number of sunspots is recorded. Note:
  - apparent periodicity.
  - large variability when series at high level; small variability when at low level.
- Series likely *stationary*.
- May have a nearly perfectly periodic component.

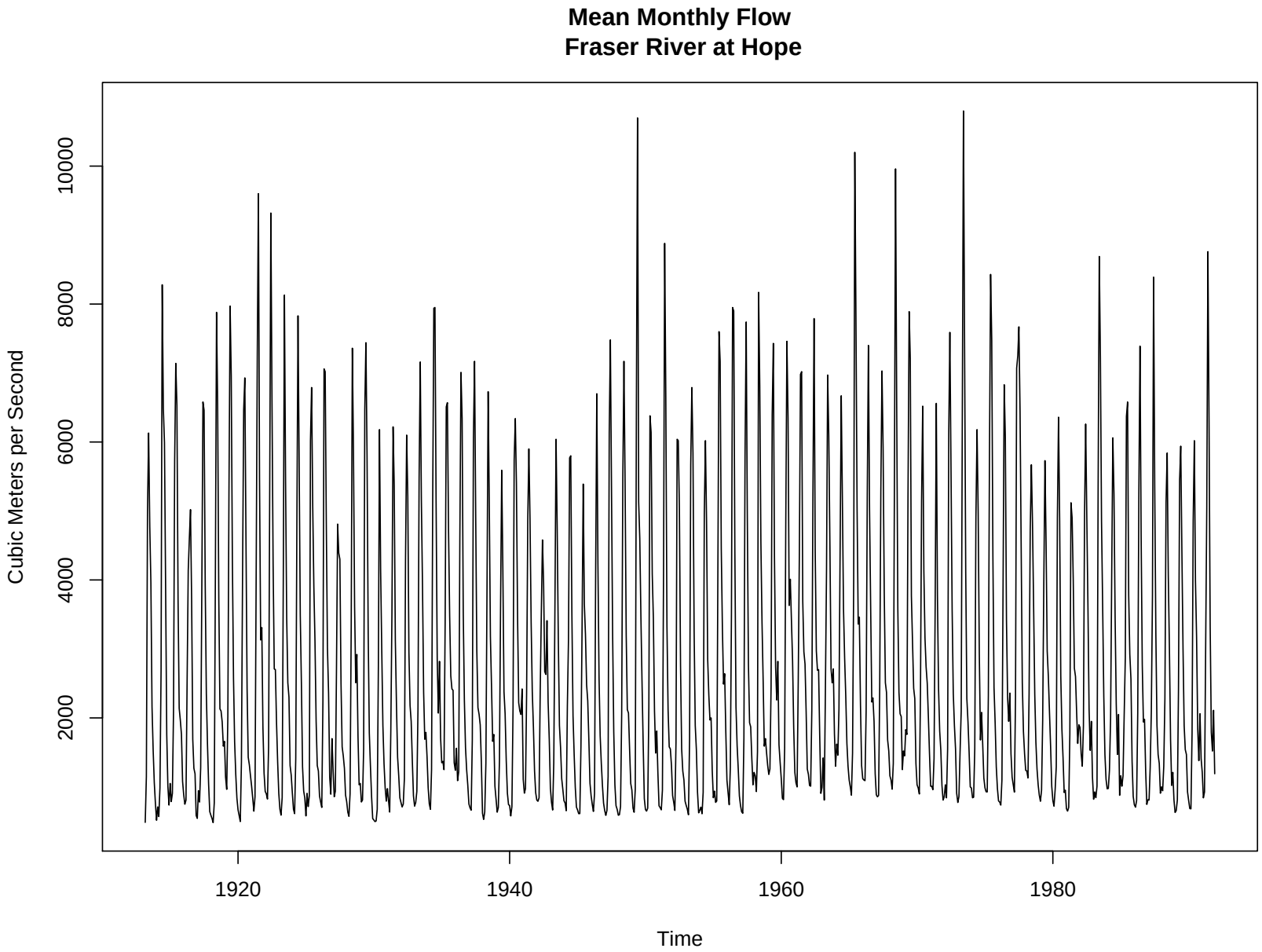
Lynx


Annual Sales of Lynx
to Hudson's Bay Co.
lynx
0
1000
2000
3000
4000
5000
6000
7000
1820
1840
1860
1880
1900
1920
Time

Comments on lynx pelt data

- Annual sales of lynx pelts to the Hudson's Bay Company. Note:
  - Clear cycle of about 10 years in length.
  - Longer term cycle?
  - Is the cycle produced by a strictly periodic phenomenon or by a dynamic system close to a periodic system?

Fraser River



Comments on Fraser River data

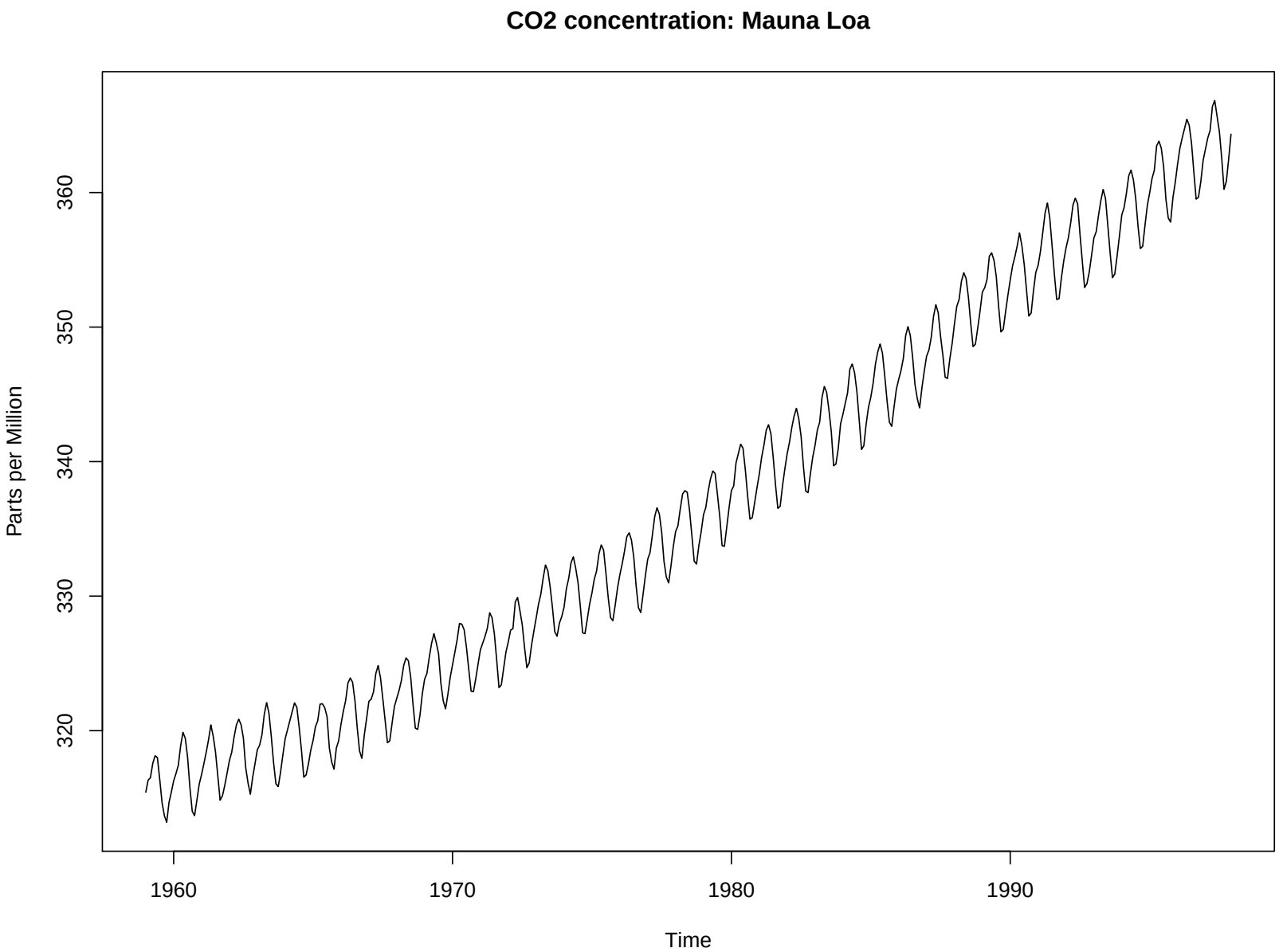
- Mean monthly flow rates for the Fraser River at Hope. Note:
  - Signs of lower variability at low levels suggesting transformation.
  - Clear annual cycle which will have to be removed to look for stationary residuals.

Canadian Unemployment

Comments on Canadian unemployment data

- Canadian monthly unemployment number. Note:
  - probable presence of slow upward trend; such a trend should be present in the presence of a growing population.
  - not stationary.
  - trend not too linear with some apparent long term cycles perhaps which produce an S shaped curve.

Mauna Loa Carbon Dioxide



Comments on Mauna Loa carbon dioxide data

- Carbon Dioxide above Mauna Loa (a Hawaiian volcano). Note:
  - Clear trend and an annual cycle
  - but you might well hope that after compensating for these the remainder would be stationary.

Changes in length of Earth's Day

Comments on earth day length data

- Changes in the length of the Earth's day. Note:
  - very smooth graph with long runs going up and down
  - suggests *integration*.

  We will look at differencing as a method of producing a series with less long range dependence.

Computing

Plots made with R using following code: # sunspots data set built into R and SPlus
lynx data set built into R and SPlus
flow monthly flows of the Fraser River
at Hope, BC.
unemployment is number unemployed in Canada
from Table 64.1: Data by Andrews and Herzberg.

Code continued

```
# changes: annual change in length of earth day
# measured in 0.00001 seconds from Table 20.1
# in Data by Andrews and Herzberg. x = read.table("table64.1",header=F)
unemployment = ts(c(t(x[,-(1:4)])), start=1956, frequency=12)
```

Code continued

```
x = read.table("table20.1",header=F)
changes = ts(c(as.matrix(x)[,c(5,7,9,11)])[1:150]
        ,start=1821, frequency=1)
flow = scan("FraserRiver.dat",skip=1)
flow = ts(flow,start=c(1913,3),frequency=12)
```

Code continued

```
par(mfrow=c(3,2))
plot(sunspots,main="Mean Monthly Sunspot Numbers") plot(lynx,
main="Annual Sales of Lynx\n to Hudson's Bay Co.")
plot(flow, ylab="Cubic Meters per Second",
      main="Mean Monthly Flow\n Fraser River at Hope")
plot(unemployment, main="Unemployment: Canada",ylab="Thousands")
plot(co2, main="CO2 concentration: Mauna Loa",
      ylab="Parts per Million")
plot(changes,
      main="Changes in length of day", ylab="0.00001 Seconds")
dev.off()
```

Mathematical Framework

- Data is a sequence of numbers; usually $y_1, \ldots, y_T$
- T is number of time units.
- Values usually evenly spaced in time.
- Call it a *discrete time* time series; not *continuous time*.
- Plots are *scatter plots* of t (on x-axis) vs y_t.
- On plots try to use real values of time t – dates if that makes sense.
- Dots are connected by straight line segments.
- Join $(1, y_1)$ to $(2, y_2)$ then to $(3, y_3)$ and so on.

Probability Models

- Treat y_ts as observed values of sequence of random variables
- Notation $\ldots, Y_{-1}, Y_0, Y_1, \ldots$.
- Or $Y_1, \ldots, Y_T$ or $\{Y_t; t \in \mathcal{T}\}$.
- Looks like notation for random sample.
- But variables usually *not independent*.
- Key question: does collecting more data help?