Time Series and Dynamic Learning

Concepts, Notation, Visualization

Ricardo Ehlers
ehlers@icmc.usp.br

Departamento de Matemática Aplicada e Estatística
Universidade de São Paulo

Models and Inference

A model is a simplification of reality
(and some are useful)

Observable quantities	Unobservable quantities
(can be measured)	(parameters / latent variables)

Approaches: Classical / Bayesian

Data: the observed values of the observable quantities.

Computation

"A big computer, a complex algorithm and a long time does not equal science." Robert Gentleman.

Aproximation

"Far better an approximate answer to the right question than the exact answer to the wrong question." John Tukey.

Models

"All models are wrong, but some are useful." George Box.

Forecasts

"Forecasts are always wrong (but we need them anyway)." Rob Hyndman

Introduction

- What is a time series?
- What is time series analysis?
- What are the objectives in time series analysis?

Time series data

A *time series* is a collection of observations taken sequentially along time.

Time series process

A *stochastic process* is a collection of random variables indexed by time.

- The most importante feature of this type of data is that neighbouring observations are dependent and we are interested in analysing and modeling these dependencies.
- The chronological order of the observations is crucial for the analysis.
- Time can be replaced by other variable like space, depth, etc.

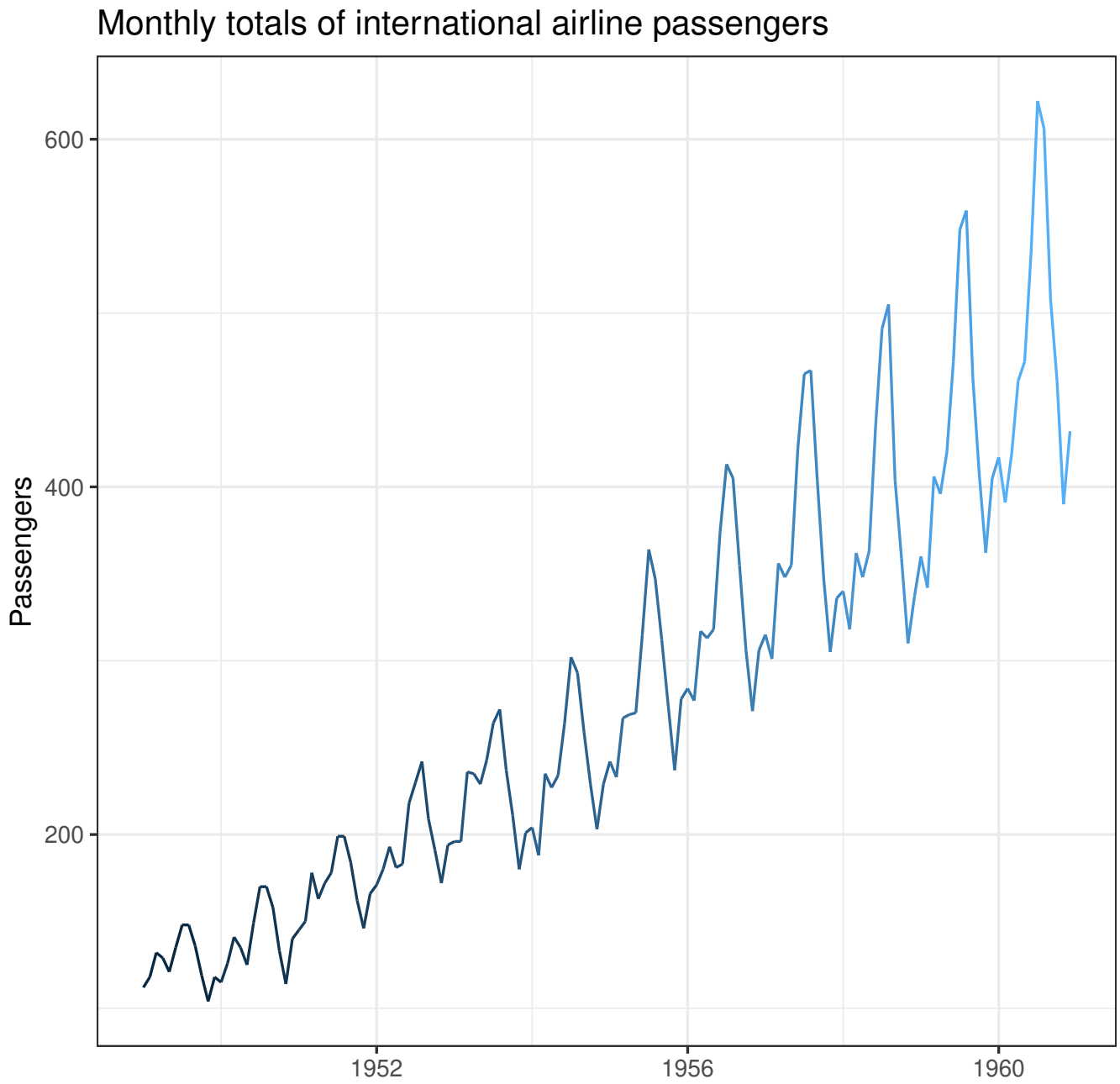
Particular features

- Correlated data are in general more difficult to analyse and require specific techniques.
- We need to take into account the chronological order of the observations.
- Complicating factors such as presence of trends and seasonal or cyclical variation can be difficult to estimate or remove.
- Model selection and comparison can be more challenging.
- It is in general more difficult to deal with missing values and outliers due to the sequential nature of the observations.
- A graphical representation of the time series is always a good starting point.

Time Series Vizualization

The ggplot2 package provides great features for time series visualization

```
> library(ggplot2)
```

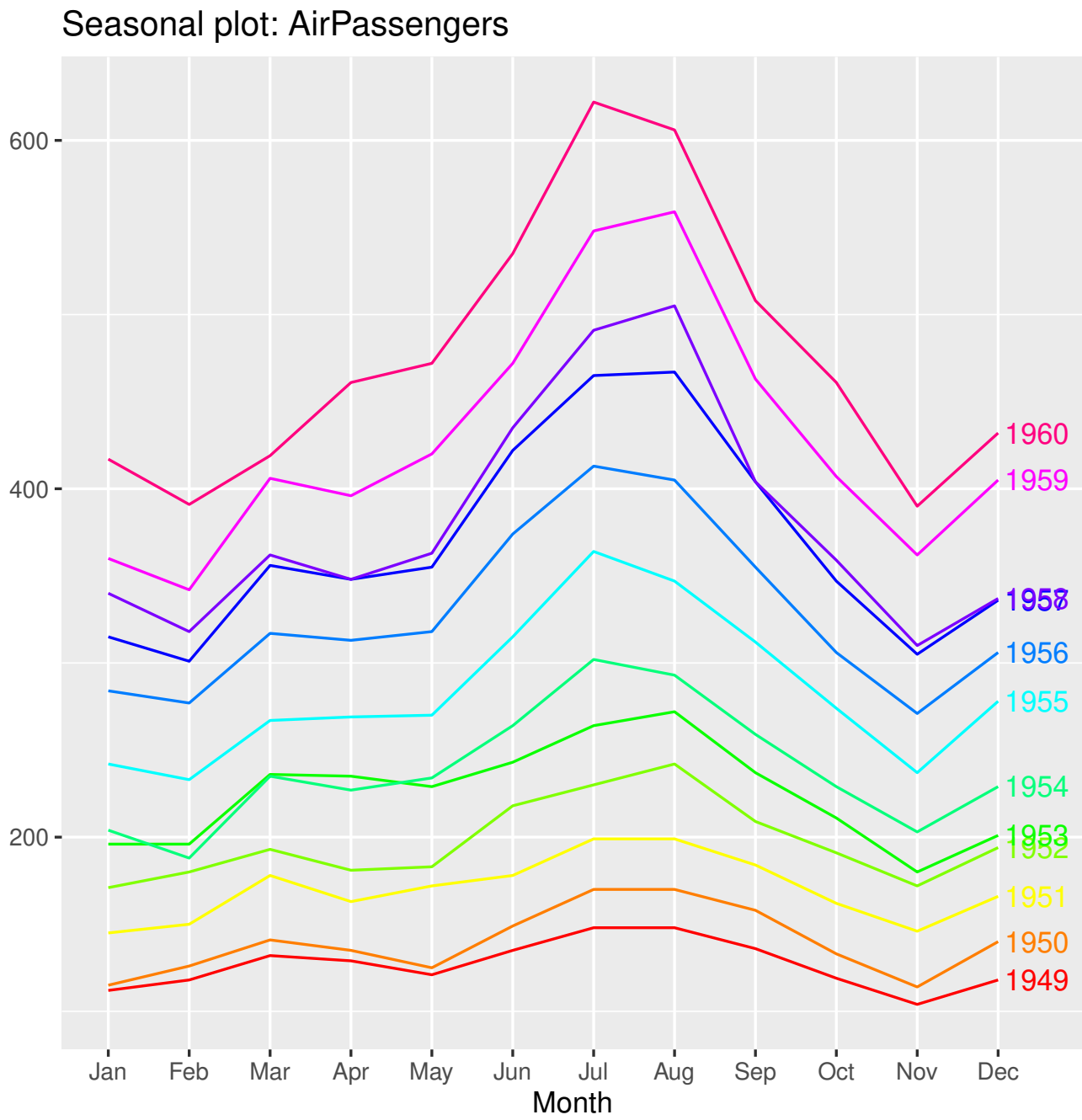

Monthly totals of international airline passengers
Passengers
600
400
200
1952
1956
1960

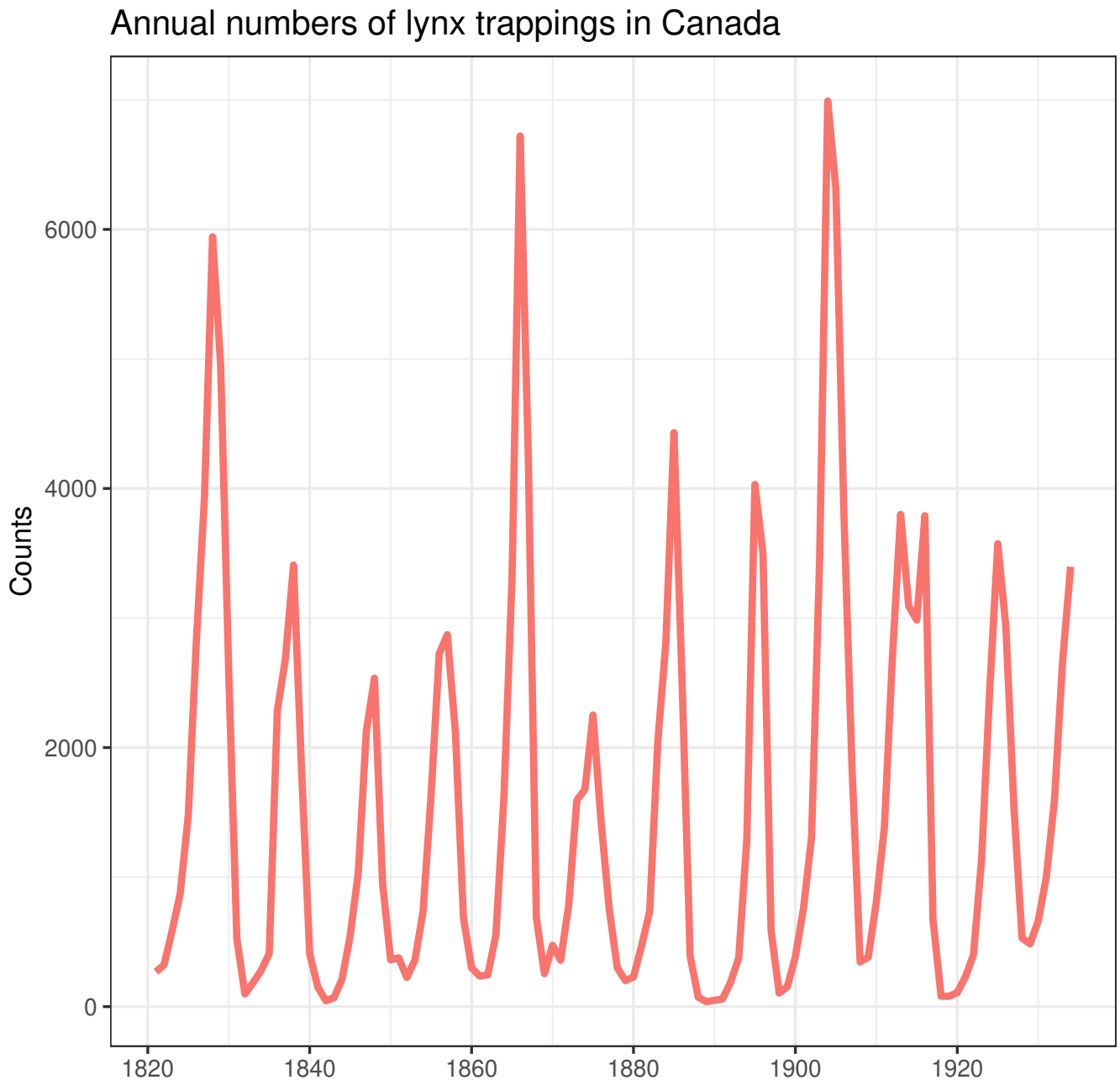
```
> help(AirPassengers)

AirPassengers              package:datasets              R Documentation

Monthly Airline Passenger Numbers 1949-1960

Description:

     The classic Box & Jenkins airline data.  Monthly totals of
     international airline passengers, 1949 to 1960.

Usage:

     AirPassengers

Format:

     A monthly time series, in thousands.

Source:

     Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (1976) _Time
     Series Analysis, Forecasting and Control._ Third Edition.
     Holden-Day. Series G.
```

```
> AirPassengers
```

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1949	112	118	132	129	121	135	148	148	136	119	104	118
1950	115	126	141	135	125	149	170	170	158	133	114	140
1951	145	150	178	163	172	178	199	199	184	162	146	166
1952	171	180	193	181	183	218	230	242	209	191	172	194
1953	196	196	236	235	229	243	264	272	237	211	180	201
1954	204	188	235	227	234	264	302	293	259	229	203	229
1955	242	233	267	269	270	315	364	347	312	274	237	278
1956	284	277	317	313	318	374	413	405	355	306	271	306
1957	315	301	356	348	355	422	465	467	404	347	305	336
1958	340	318	362	348	363	435	491	505	404	359	310	337
1959	360	342	406	396	420	472	548	559	463	407	362	405
1960	417	391	419	461	472	535	622	606	508	461	390	432

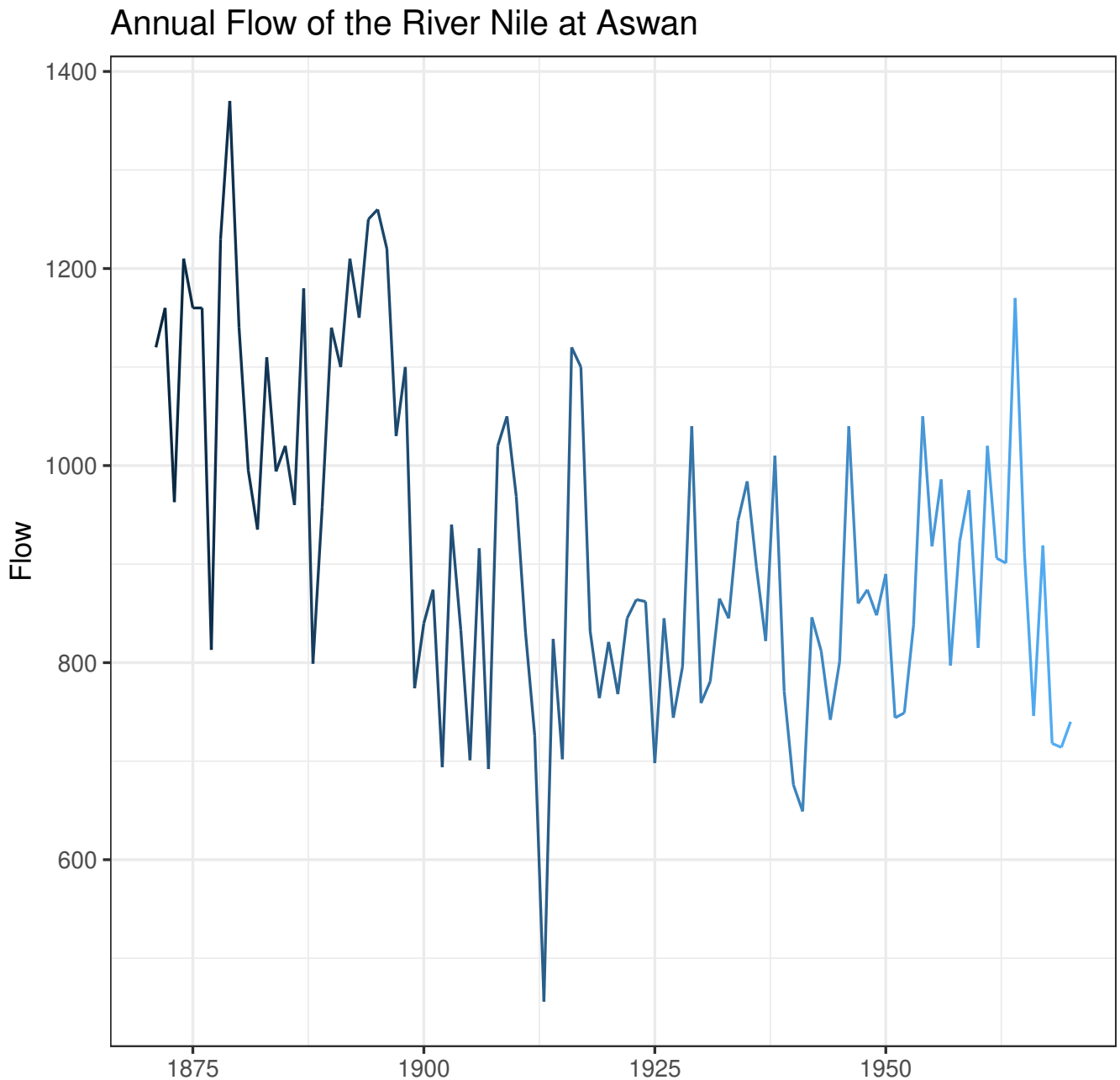
```
> library(ggplot2)
> Year= as.integer(time(AirPassengers))
> Time= as.numeric(time(AirPassengers))
> data= data.frame(Year=Year,Time=Time,
+                  Passengers=as.numeric(AirPassengers))
> ggplot(data,aes(x=Time,y=Passengers,color=Year)) +
+ geom_line() +
+ theme_bw() +
+ theme(legend.position = "none") +
+ ggtitle("Monthly totals of international airline passengers") +
+ labs(x="")
```


Seasonal plot: AirPassengers
600
400
200
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month
1960
1959
1957
1956
1955
1954
1953
1952
1951
1950
1949

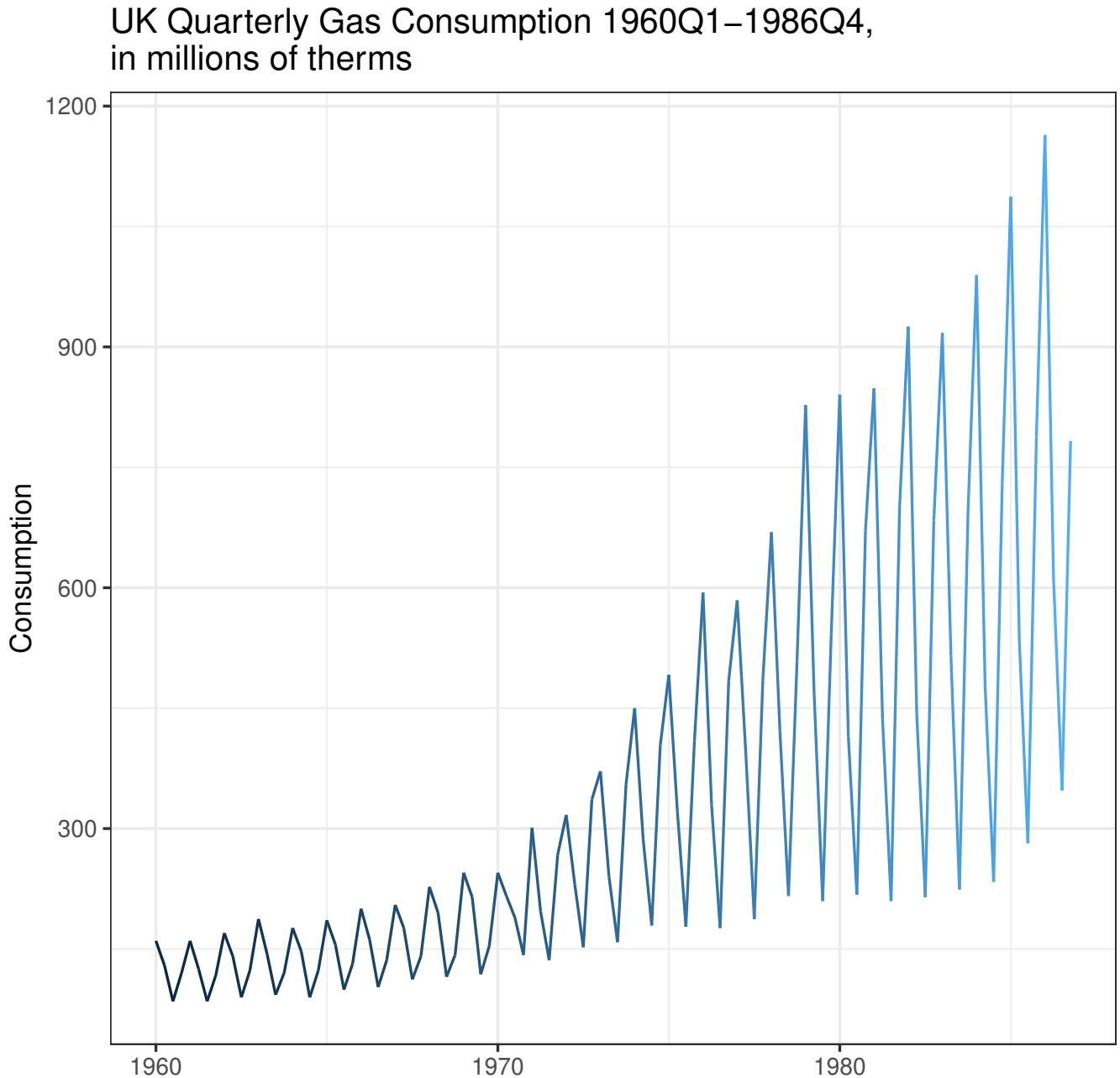

Annual numbers of lynx trappings in Canada
Counts
6000
4000
2000
0
1820
1840
1860
1880
1900
1920

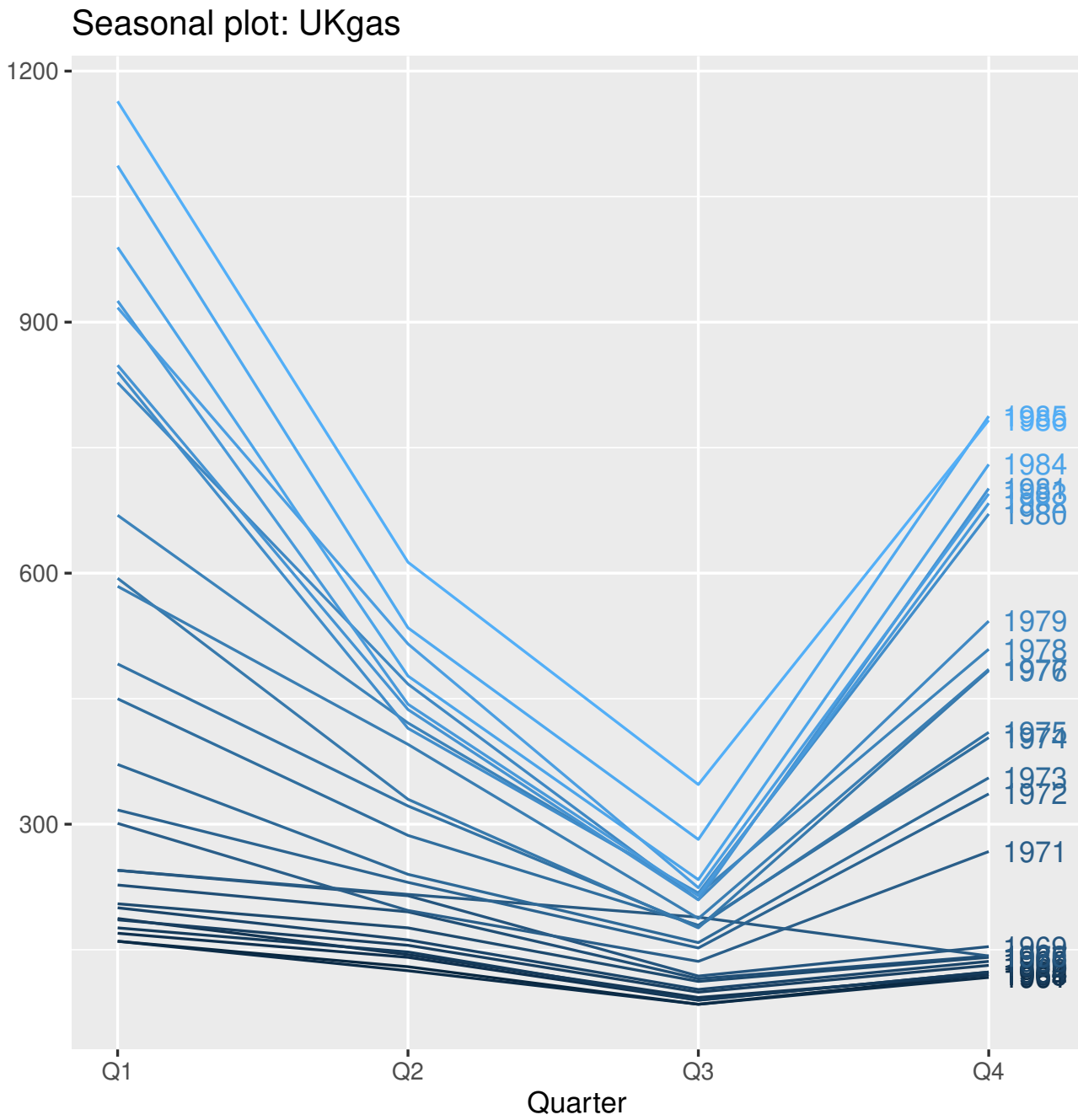
```
> help(lynx)

lynx                    package:datasets                    R Documentation

Annual Canadian Lynx trappings 1821-1934

Description:

     Annual numbers of lynx trappings for 1821-1934 in Canada. Taken
     from Brockwell & Davis (1991), this appears to be the series
     considered by Campbell & Walker (1977).

Usage:

     lynx

Source:

     Brockwell, P. J. and Davis, R. A. (1991). _Time Series and
     Forecasting Methods_. Second edition. Springer. Series G (page
     557).

References:

     Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988). _The New
     S Language_. Wadsworth & Brooks/Cole.

     Campbell, M. J. and Walker, A. M. (1977). A Survey of statistical
     work on the Mackenzie River series of annual Canadian lynx
     trappings for the years 1821-1934 and a new analysis. Journal of
     the Royal Statistical Society series A, *140*, 411-431. doi:
     10.2307/2345277 (https://doi.org/10.2307/2345277).
```

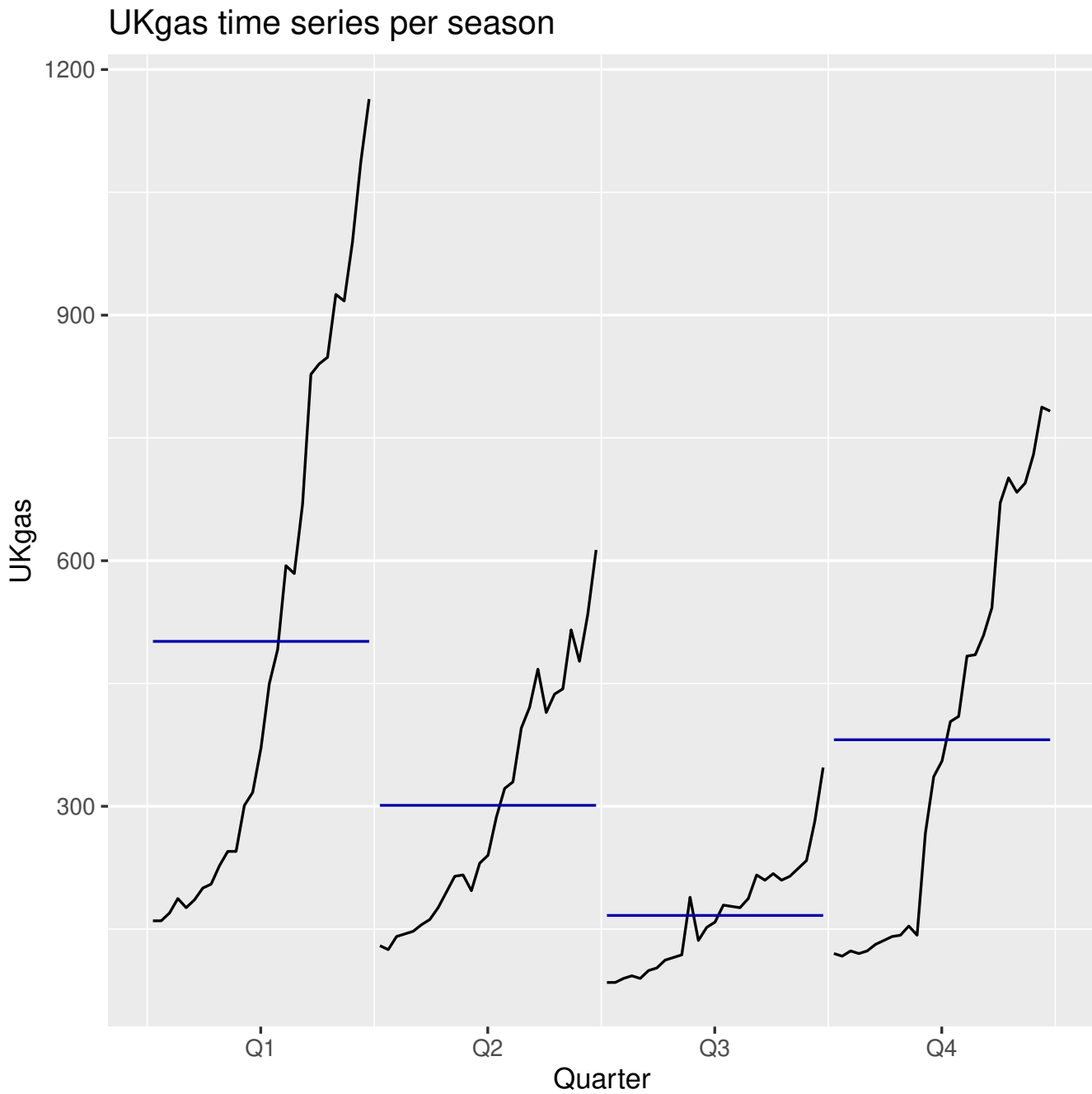
```
> lynx

Time Series:
Start = 1821
End = 1934
Frequency = 1
  [1]  269  321  585  871 1475 2821 3928 5943 4950 2577  523
 [12]   98  184  279  409 2285 2685 3409 1824  409  151   45
 [23]   68  213  546 1033 2129 2536  957  361  377  225  360
 [34]  731 1638 2725 2871 2119  684  299  236  245  552 1623
 [45] 3311 6721 4254  687  255  473  358  784 1594 1676 2251
 [56] 1426  756  299  201  229  469  736 2042 2811 4431 2511
 [67]  389   73   39   49   59  188  377 1292 4031 3495  587
 [78]  105  153  387  758 1307 3465 6991 6313 3794 1836  345
 [89]  382  808 1388 2713 3800 3091 2985 3790  674   81   80
[100]  108  229  399 1132 2432 3574 2935 1537  529  485  662
[111] 1000 1590 2657 3396
```

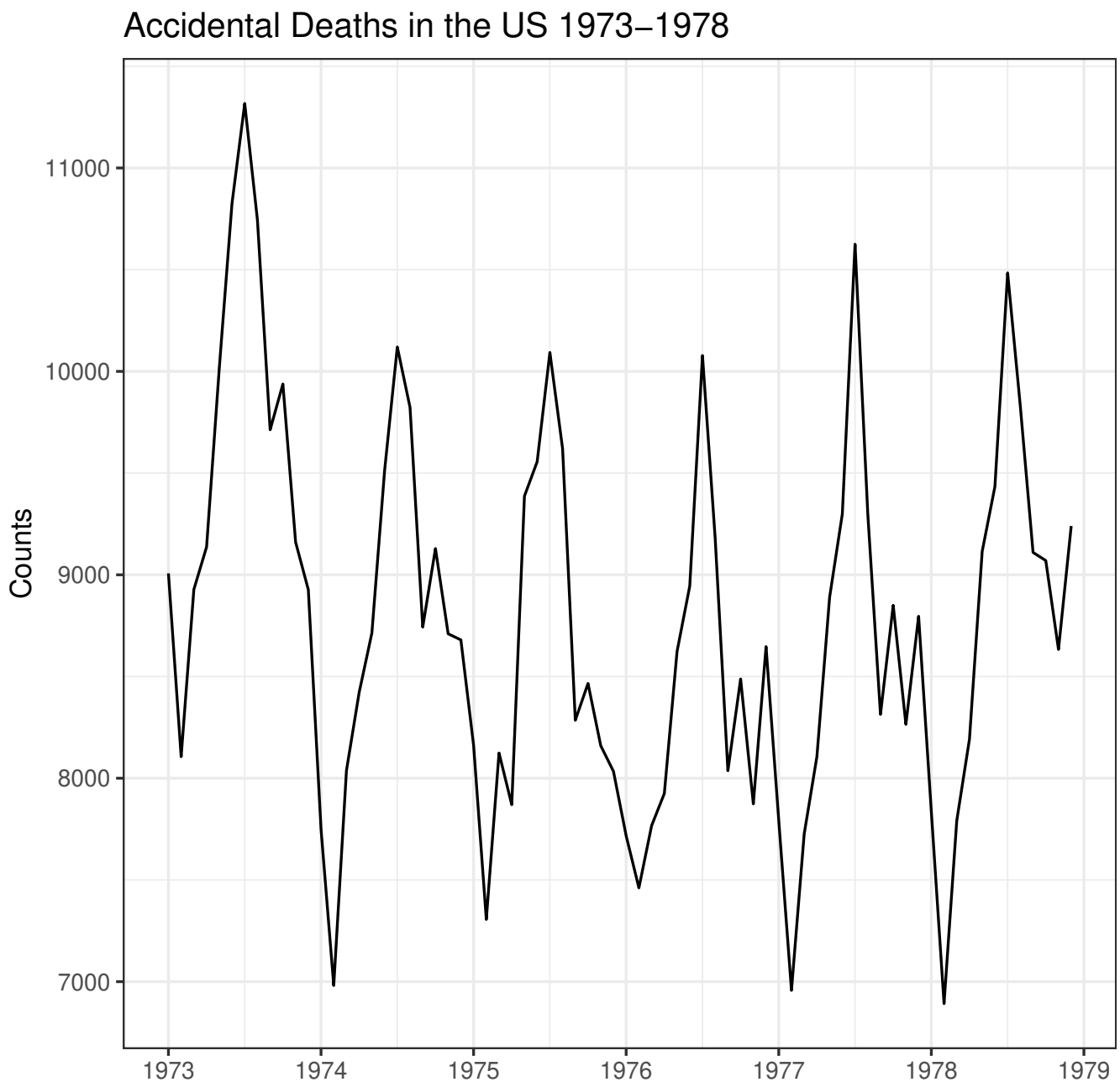

Annual Flow of the River Nile at Aswan
Flow
1400
1200
1000
800
600
1875
1900
1925
1950

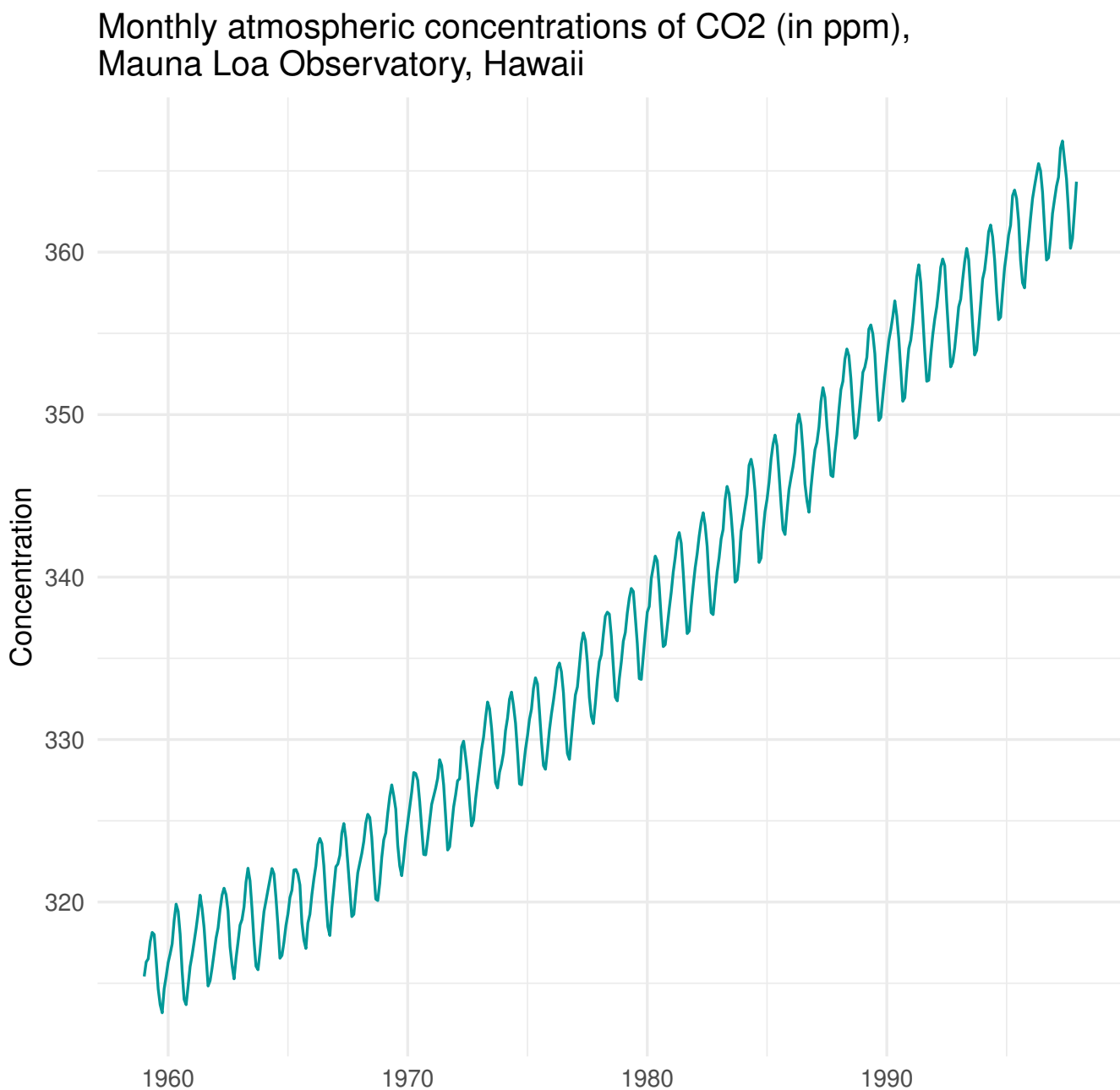
```
> help(Nile)

Nile                    package:datasets                    R Documentation

Flow of the River Nile

Description:

     Measurements of the annual flow of the river Nile at Aswan
     (formerly 'Assuan'), 1871-1970, in 10^8 m^3, "with apparent
     changepoint near 1898" (Cobb(1978), Table 1, p.249).

Usage:

     Nile

Format:

     A time series of length 100.

Source:

     Durbin, J. and Koopman, S. J. (2001).  _Time Series Analysis by
     State Space Methods_.  Oxford University Press.  <URL:
     http://www.ssfpack.com/DKbook.html>

References:

     Balke, N. S. (1993).  Detecting level shifts in time series.
     _Journal of Business and Economic Statistics_, *11*, 81-92.  doi:
     10.2307/1391308 (https://doi.org/10.2307/1391308).

     Cobb, G. W. (1978).  The problem of the Nile: conditional solution
     to a change-point problem.  _Biometrika_ *65*, 243-51.  doi:
     10.2307/2335202 (https://doi.org/10.2307/2335202).
```

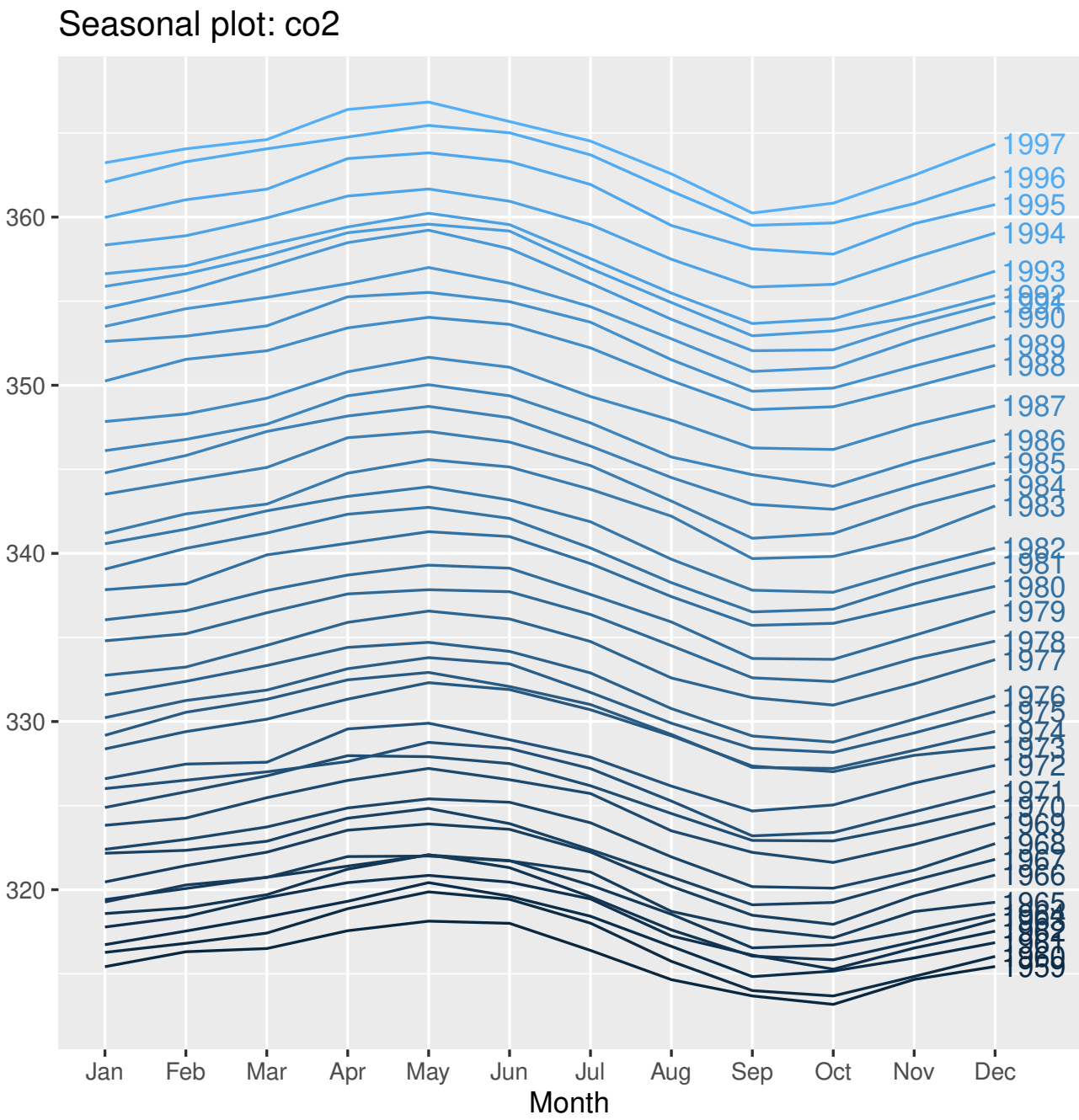

UK Quarterly Gas Consumption 1960Q1–1986Q4,
in millions of therms
Consumption
1200
900
600
300
1960
1970
1980

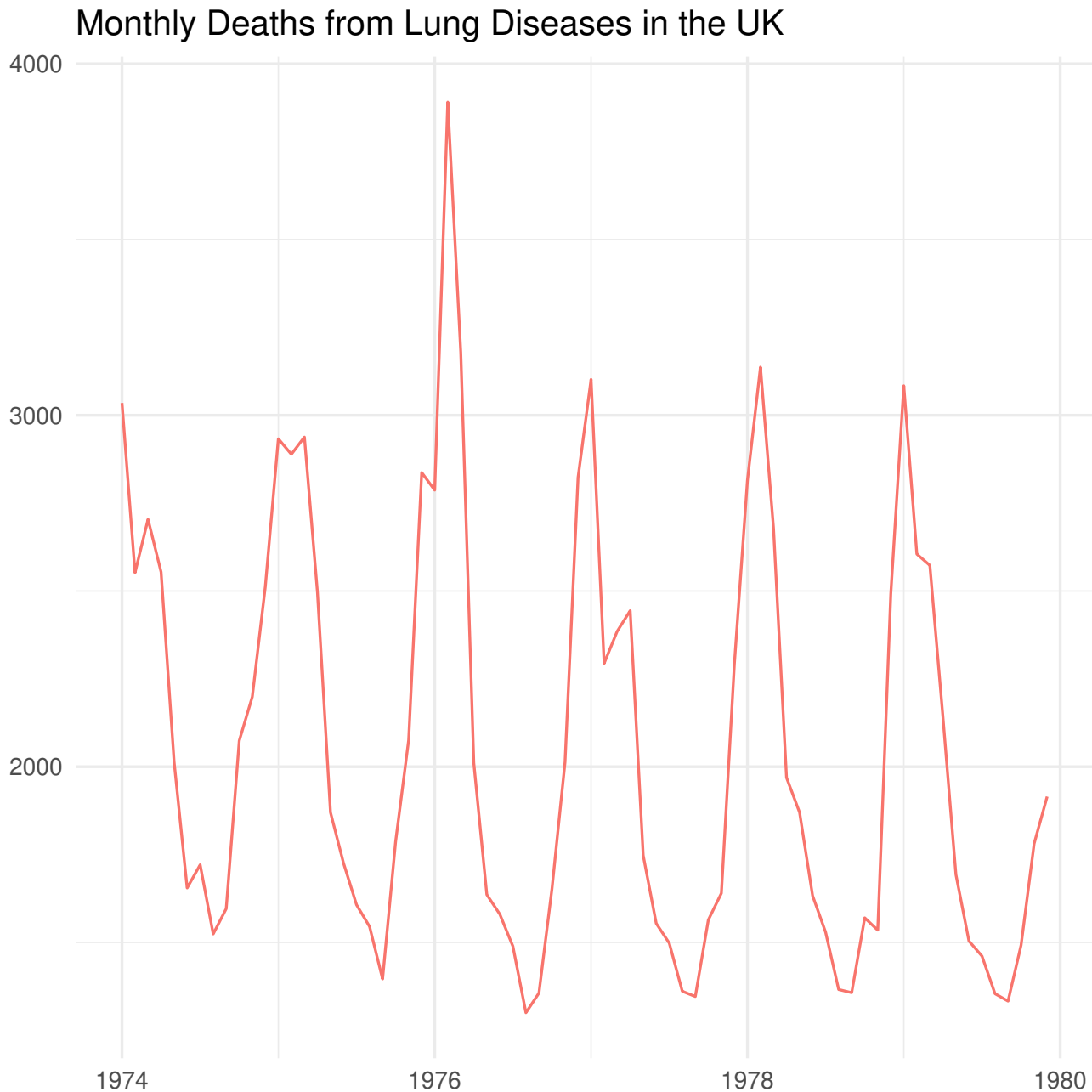

Seasonal plot: UKgas
1200
900
600
300
Q1
Q2
Q3
Q4
Quarter
1984
1980
1979
1978
1976
1974
1973
1972
1971

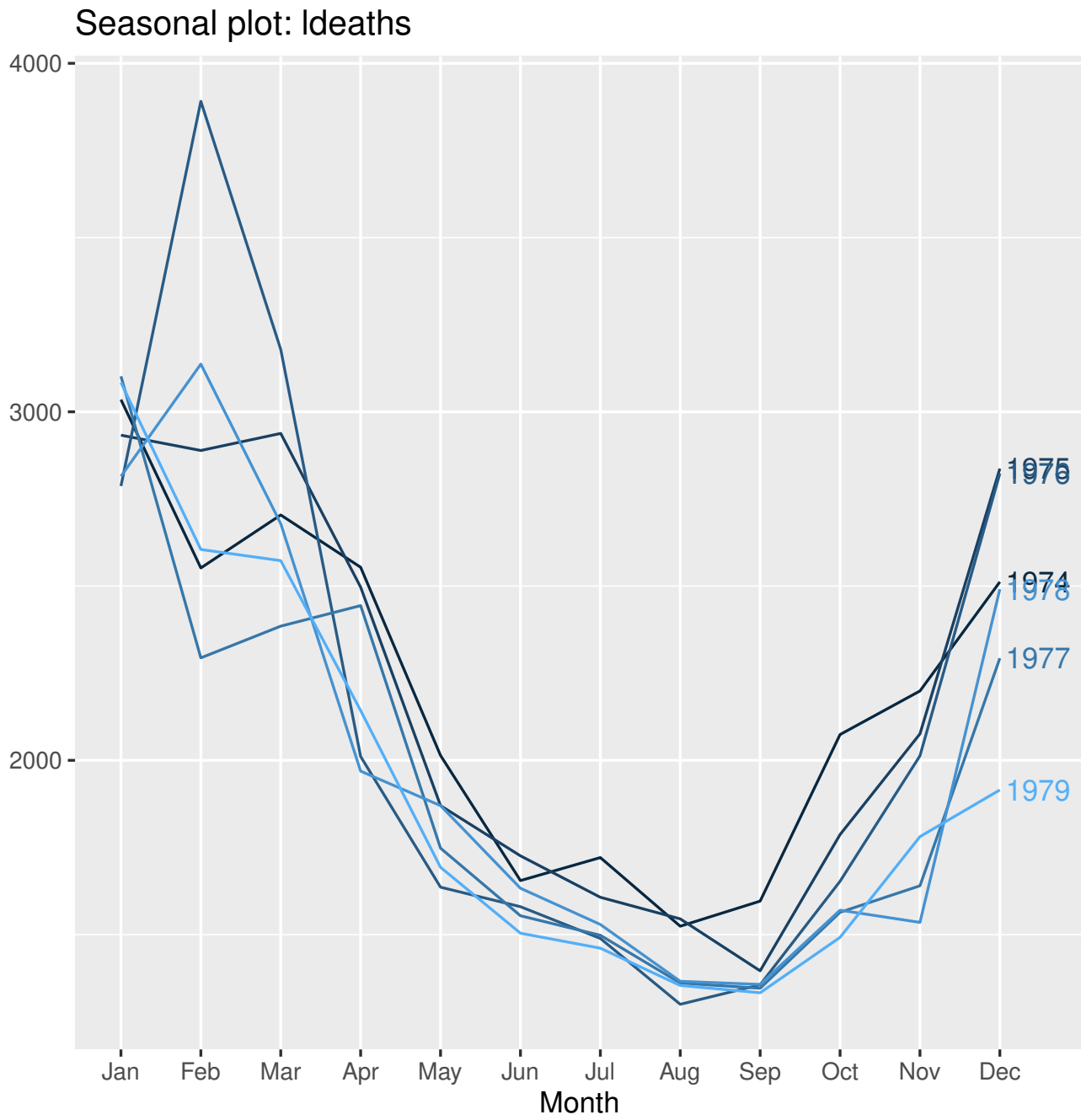

UKgas time series per season
1200
900
600
300
UKgas
Q1
Q2
Q3
Q4
Quarter


Accidental Deaths in the US 1973–1978
Counts
11000
10000
9000
8000
7000
1973
1974
1975
1976
1977
1978
1979

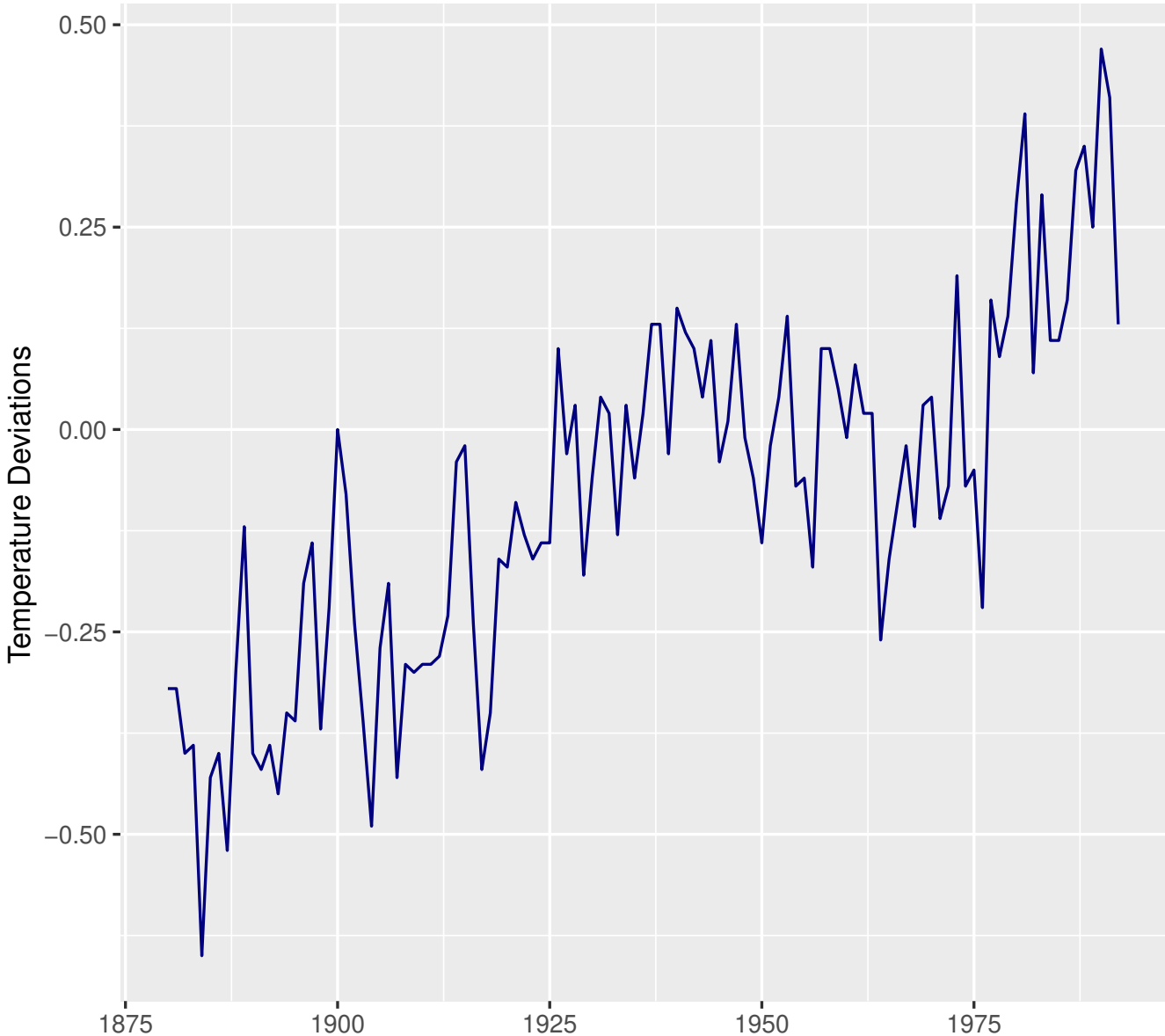

Monthly atmospheric concentrations of CO2 (in ppm),
Mauna Loa Observatory, Hawaii
Concentration
360
350
340
330
320
1960
1970
1980
1990


Seasonal plot: co2
360
350
340
330
320
1997
1996
1995
1994
1993
1990
1989
1988
1987
1986
1985
1984
1983
1982
1980
1979
1978
1977
1976
1974
1973
1972
1971
1966
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month

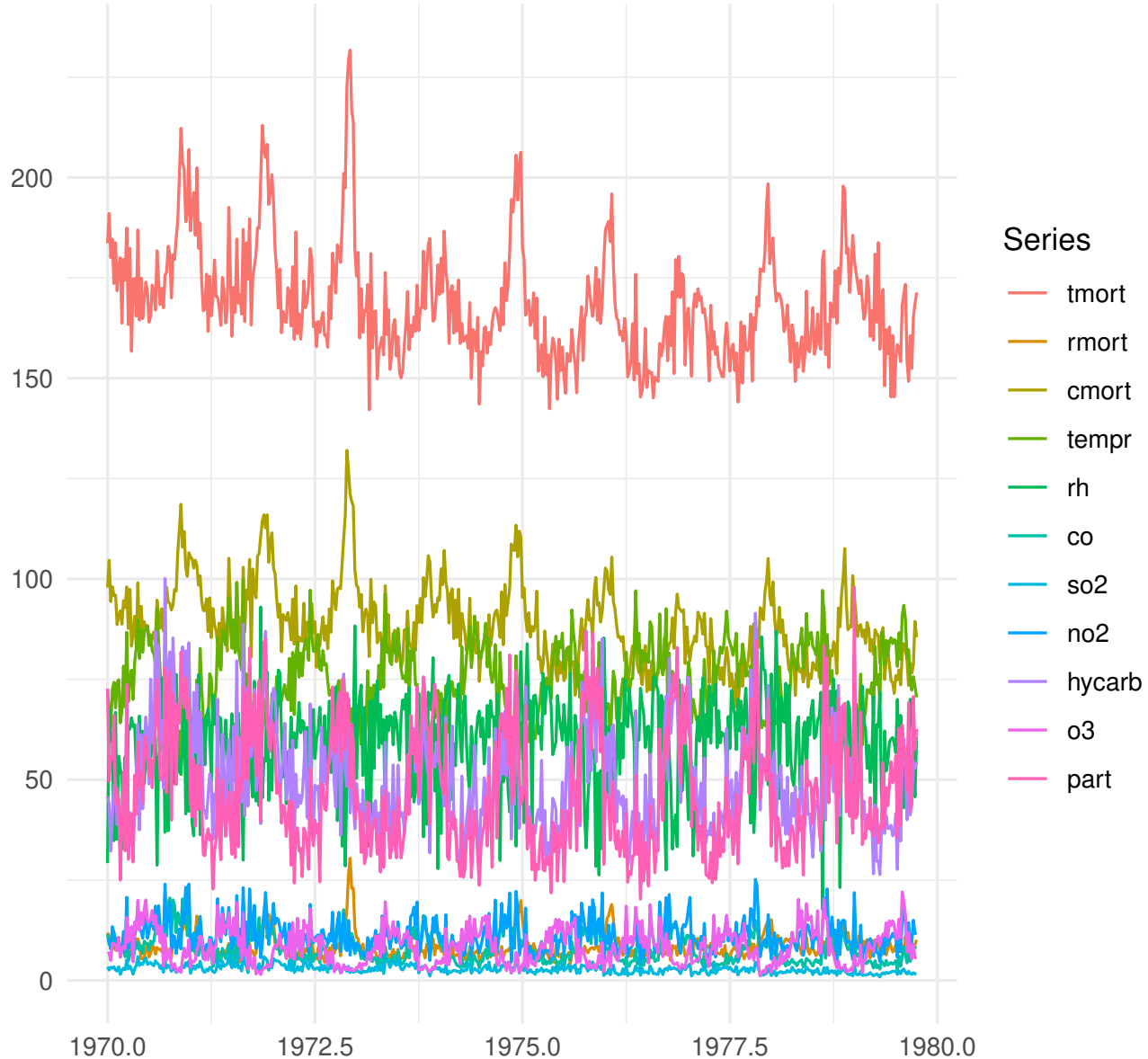

Monthly Deaths from Lung Diseases in the UK
4000
3000
2000
1974
1976
1978
1980


Seasonal plot: ldeaths
4000
3000
2000
1975
1976
1974
1978
1977
1979
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month

Global mean land–ocean temperature deviations, 1880–2015
http://data.giss.nasa.gov/gistemp/graphs

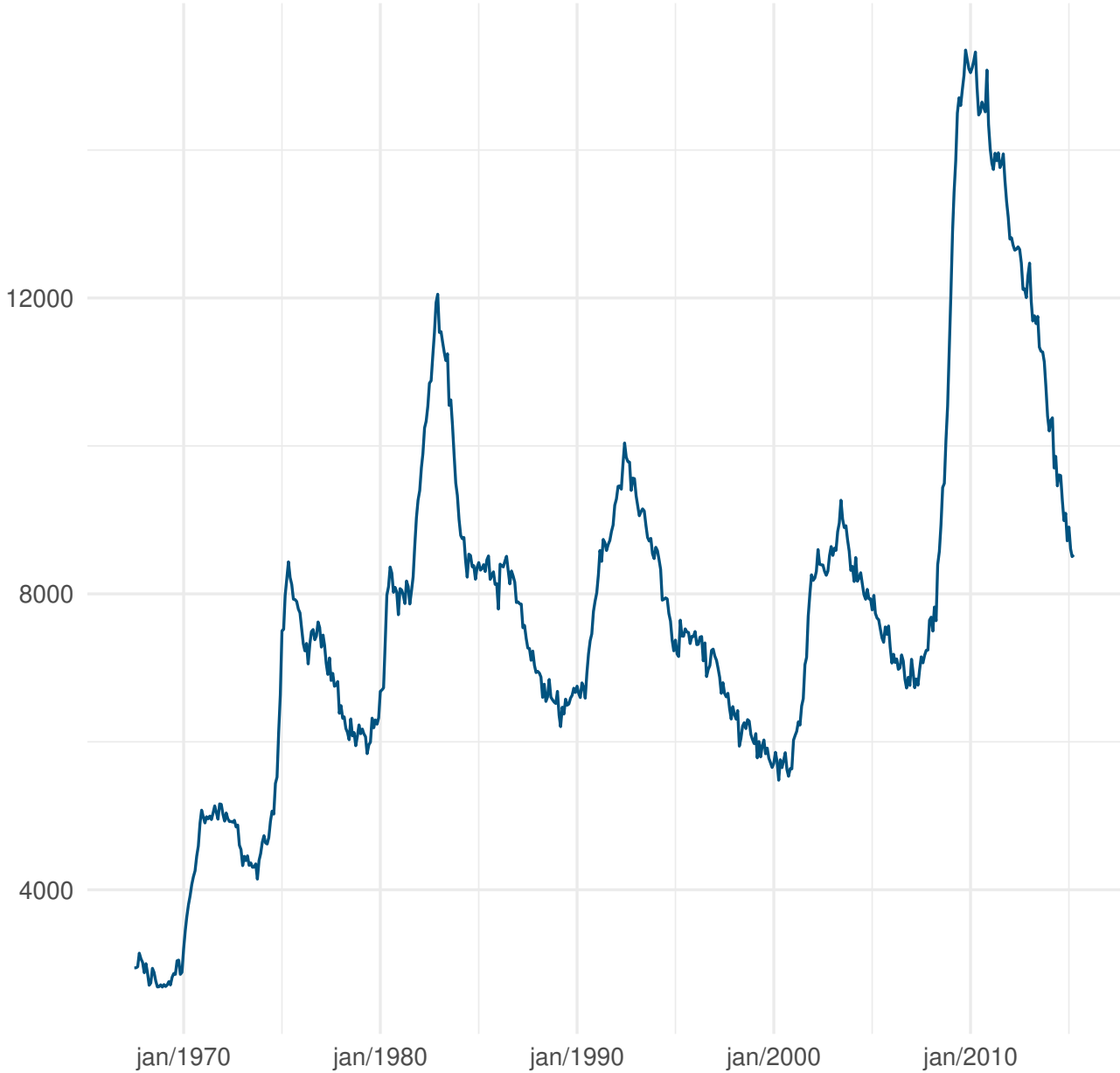


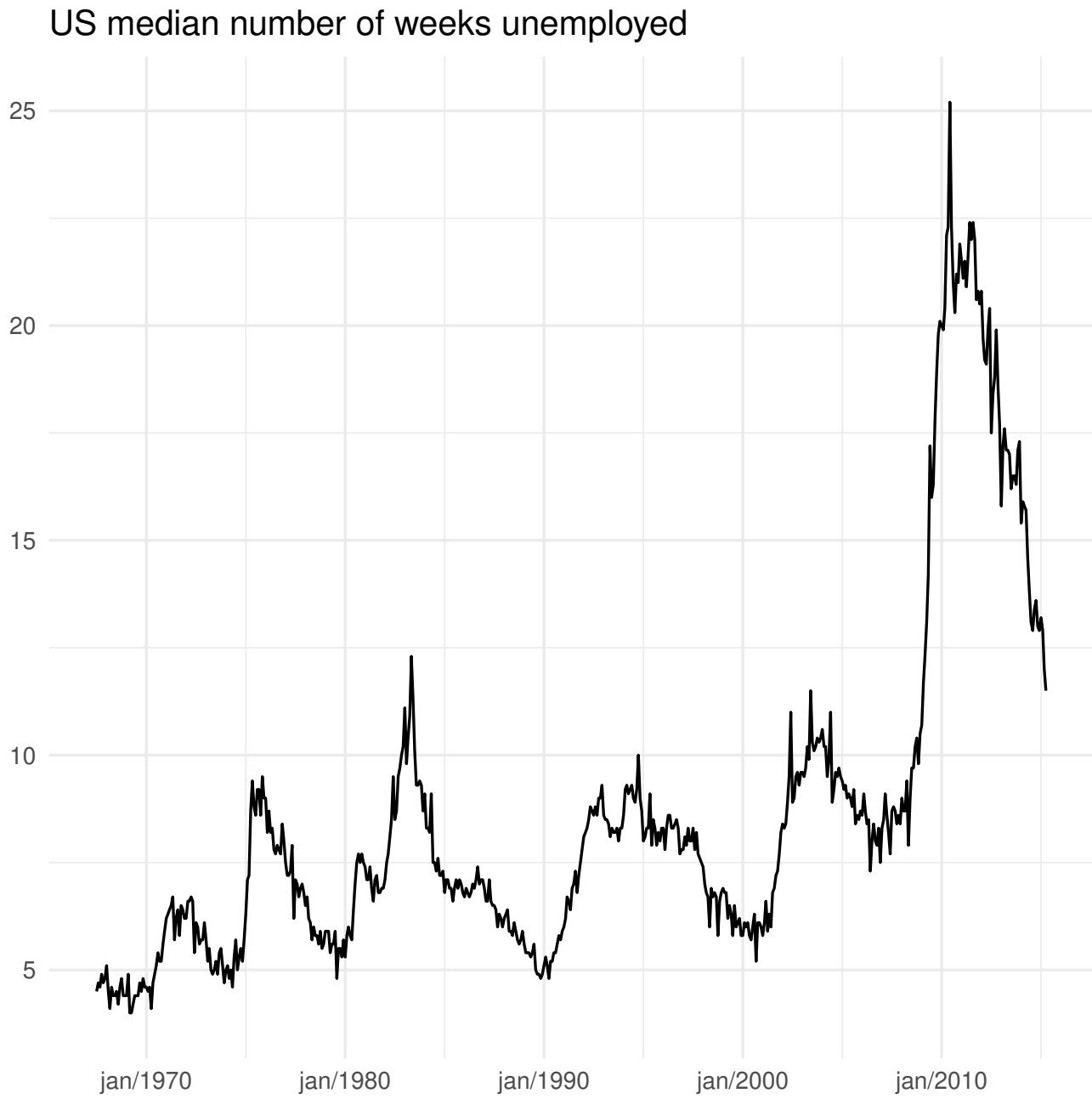
LA Pollution–Mortality Study (1970–1979, weekly data)
http://www.stat.pitt.edu/stoffer/tsda/

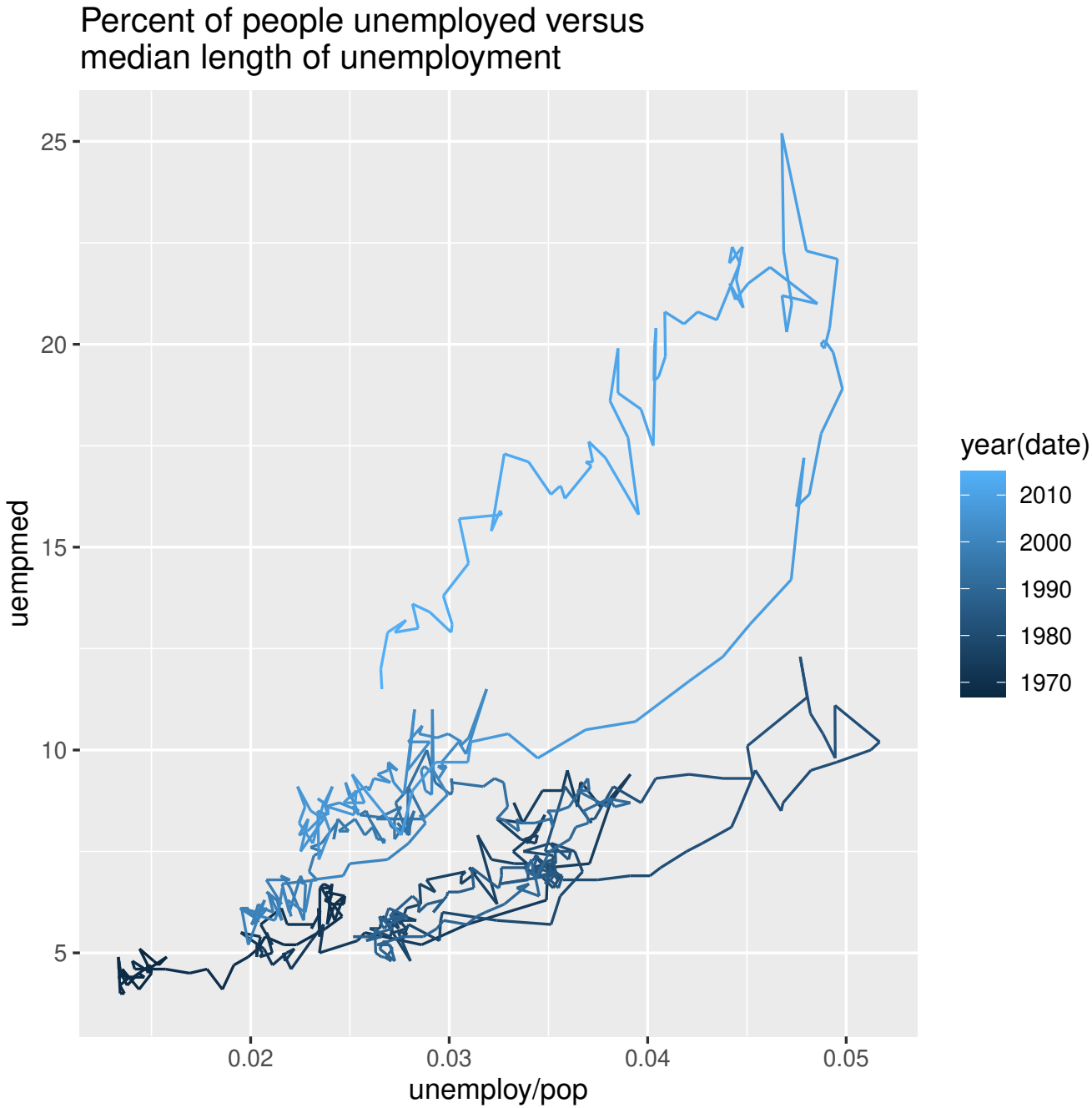


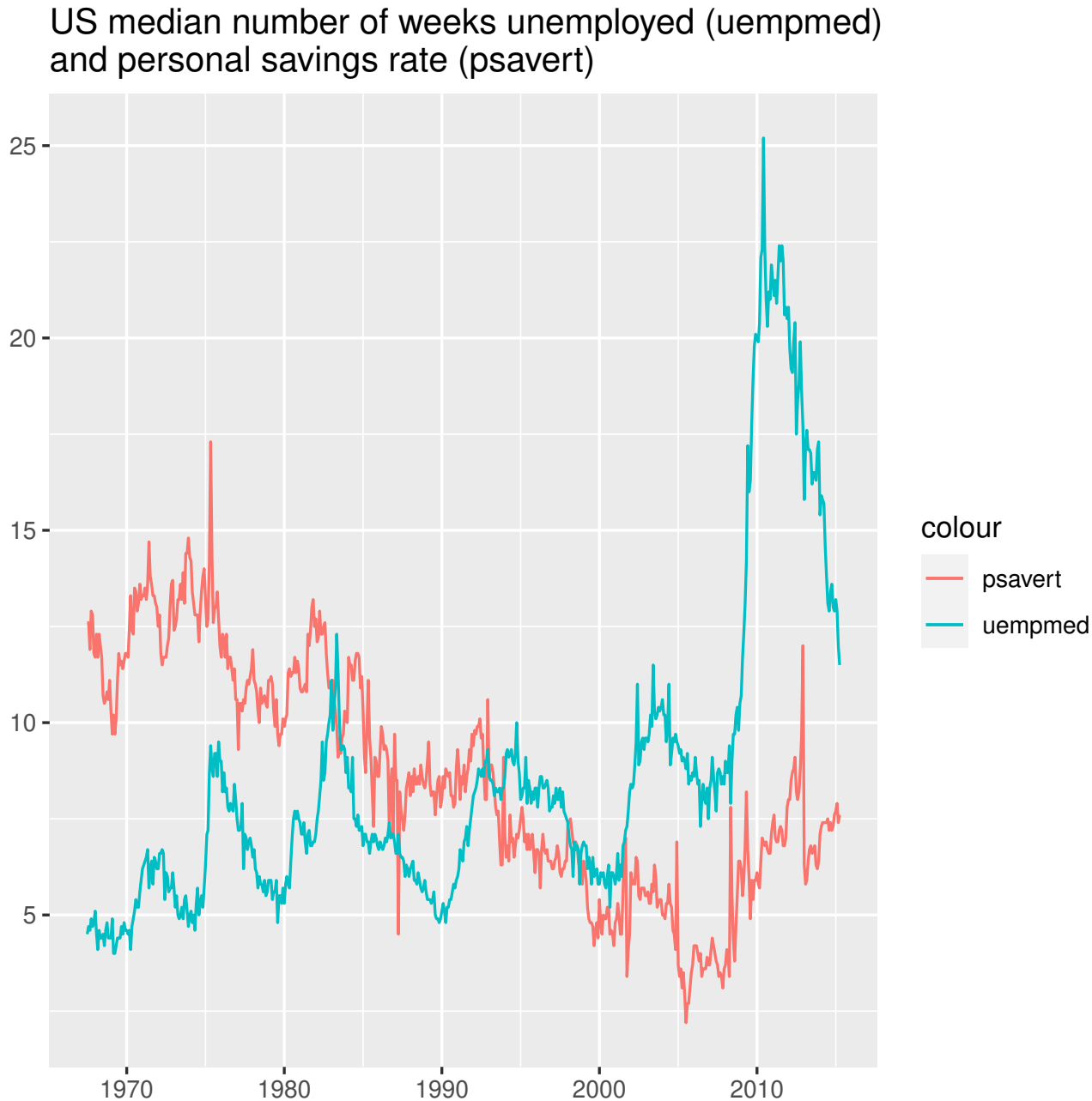
```
> help(lap)

lap                    package:astsa                    R Documentation

LA Pollution-Mortality Study

Description:

     LA Pollution-Mortality Study (1970-1979, weekly data).

Format:

     The format is: mts [1:508, 1:11]

Details:

       columns are time series       with names
       (1) Total Mortality           'tmort'
       (2) Respiratory Mortality     'rmort'
       (3) Cardiovascular Mortality  'cmort'
       (4) Temperature               'tempr'
       (5) Relative Humidity         'rh'
       (6) Carbon Monoxide           'co'
       (7) Sulfur Dioxide            'so2'
       (8) Nitrogen Dioxide          'no2'
       (9) Hydrocarbons              'hycarb'
       (10) Ozone                    'o3'
       (11) Particulates             'part'

References:

    <http://www.stat.pitt.edu/stoffer/tsa4/> and <URL:
     http://www.stat.pitt.edu/stoffer/tsda/>
```

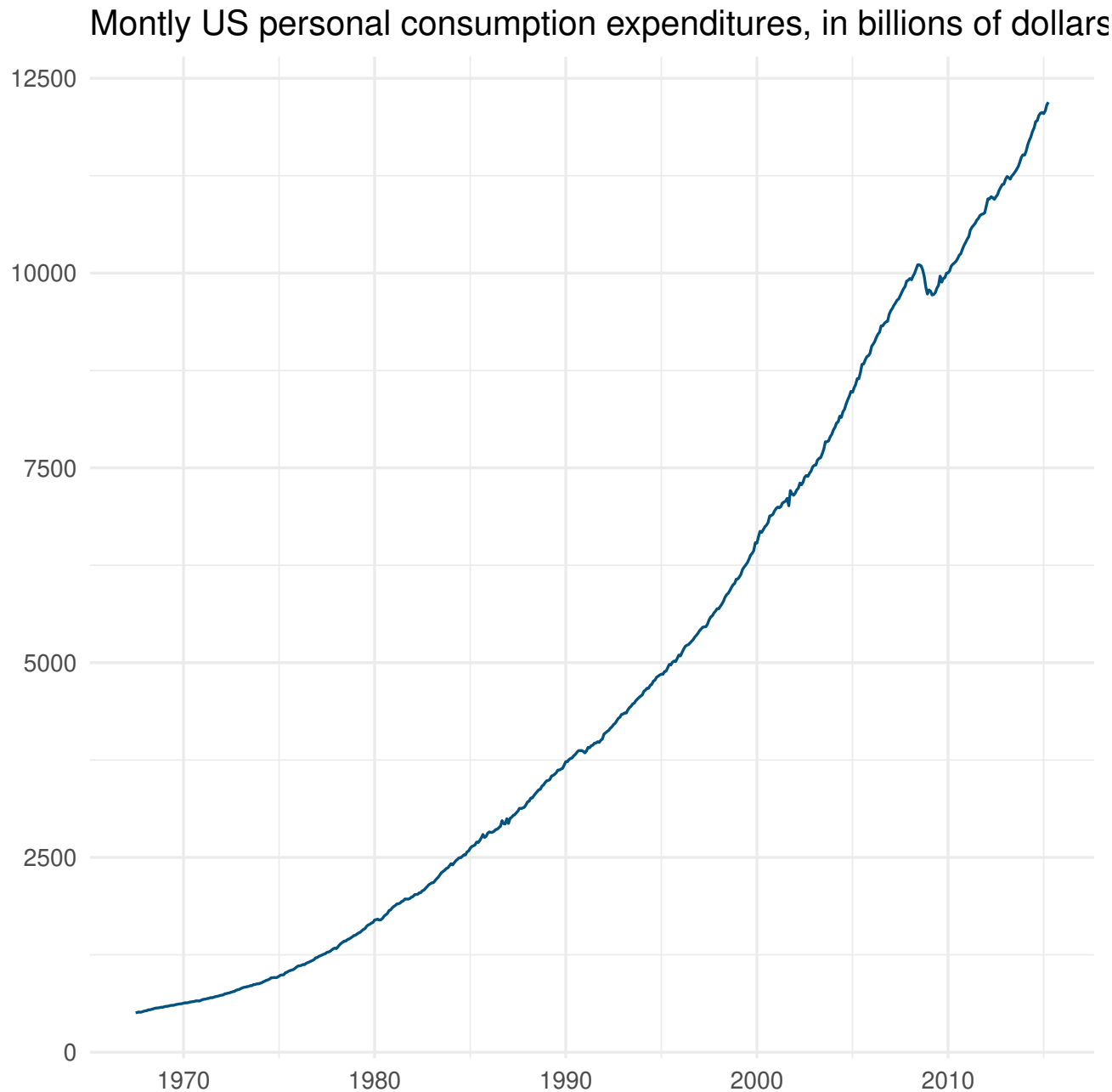
US Unemployed in Thousands (http://research.stlouisfed.org/fred2)

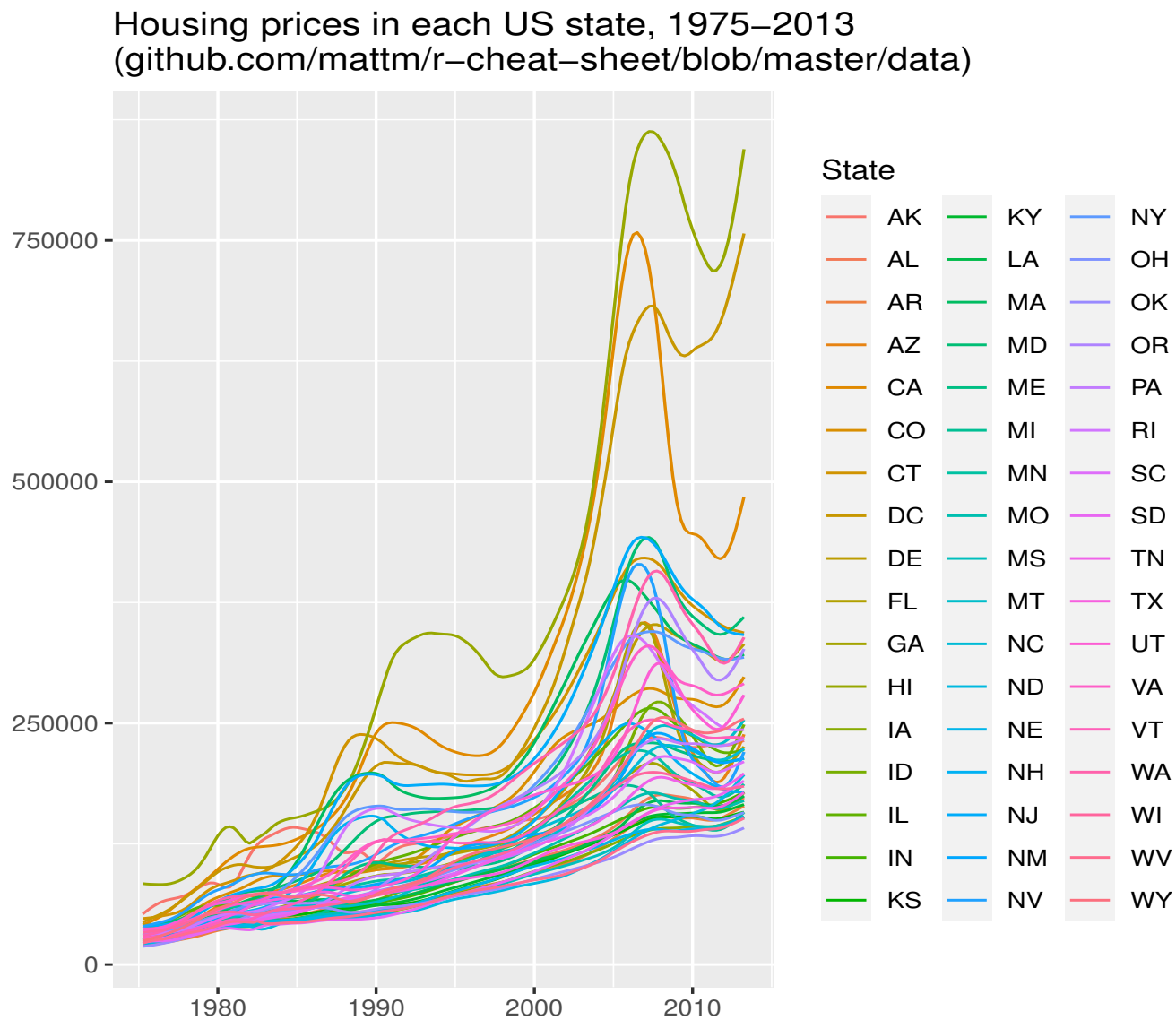


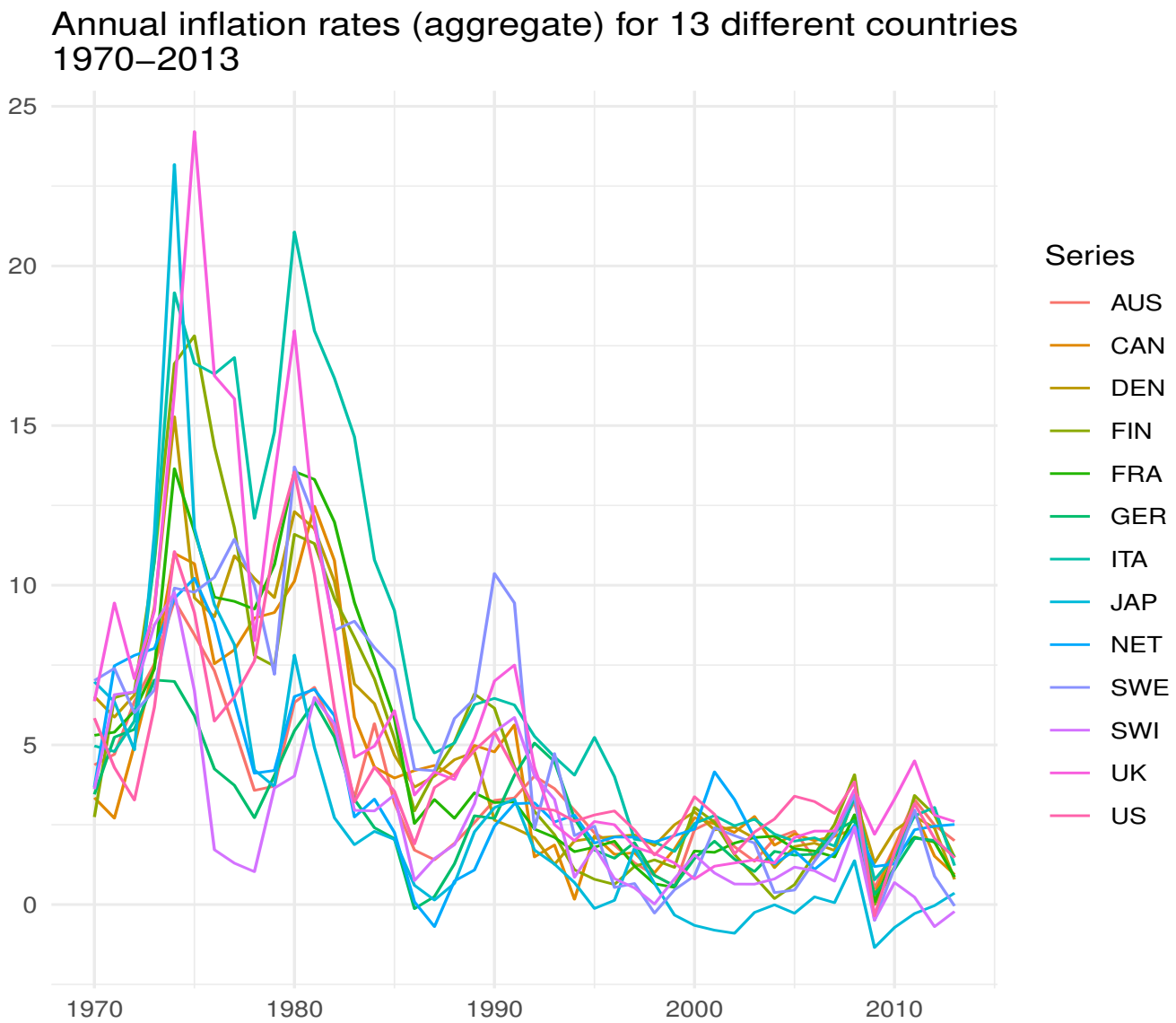

US median number of weeks unemployed
25
20
15
10
5
jan/1970
jan/1980
jan/1990
jan/2000
jan/2010

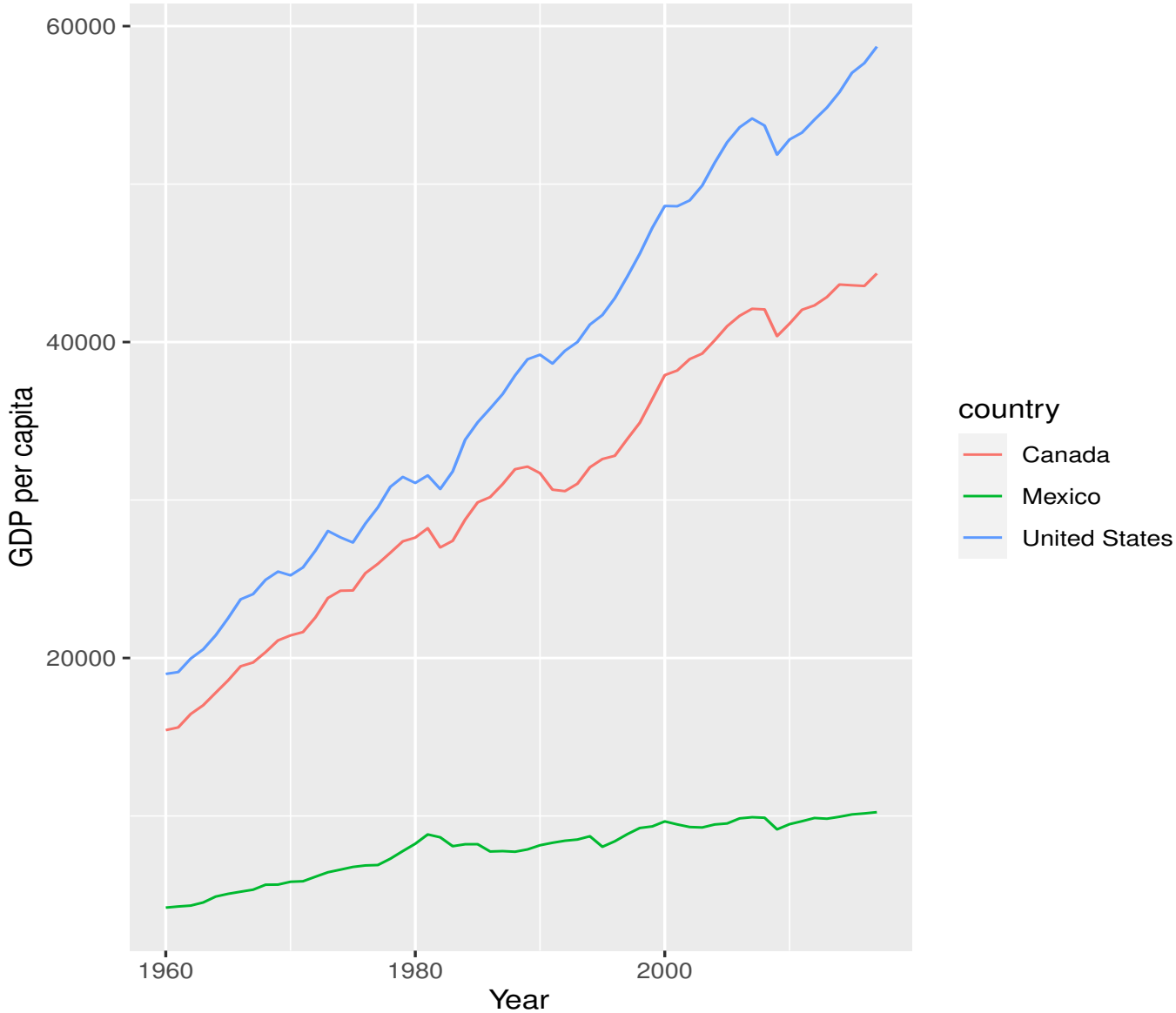

Percent of people unemployed versus
median length of unemployment
uempmed
25
20
15
10
5
0.02
0.03
0.04
0.05
unemploy/pop
year(date)
2010
2000
1990
1980
1970


US median number of weeks unemployed (uempmed)
and personal savings rate (psavert)
25
20
15
10
5
1970
1980
1990
2000
2010
colour
psavert
uempmed


Montly US personal consumption expenditures, in billions of dollars
12500
10000
7500
5000
2500
0
1970
1980
1990
2000
2010


Housing prices in each US state, 1975–2013
(github.com/mattm/r–cheat–sheet/blob/master/data)
750000
500000
250000
0
1980
1990
2000
2010
State
AK
AL
AR
AZ
CA
CO
CT
DC
DE
FL
GA
HI
IA
ID
IL
IN
KS
KY
LA
MA
MD
ME
MI
MN
MO
MS
MT
NC
ND
NE
NH
NJ
NM
NV
NY
OH
OK
OR
PA
RI
SC
SD
TN
TX
UT
VA
VT
WA
WI
WV
WY


Annual inflation rates (aggregate) for 13 different countries
1970–2013
25
20
15
10
5
0
1970
1980
1990
2000
2010
Series
AUS
CAN
DEN
FIN
FRA
GER
ITA
JAP
NET
SWE
SWI
UK
US


60000
40000
20000
GDP per capita
1960
1980
2000
Year
country
Canada
Mexico
United States

Financial Returns

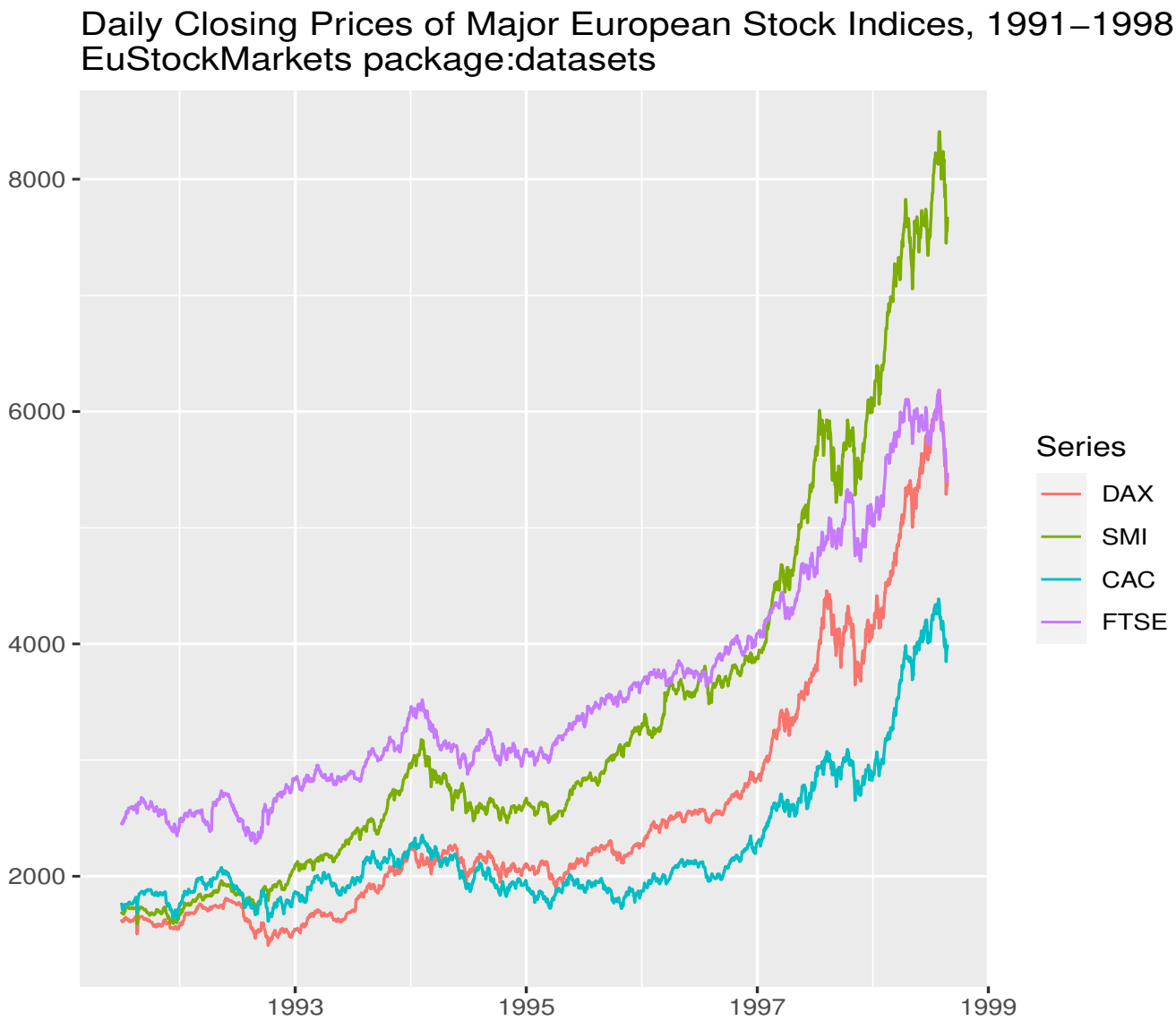
In Finance, risk is frequently measured in terms of price variations of an asset. Let P_t the price of an asset at time t.

- The price variation between $t-1$ and t (no dividends paid) is, $P_t - P_{t-1}$.
- Relative variation of prices,

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} = \frac{P_t}{P_{t-1}} - 1.$$

- Continuosly compounded return, or log-return,

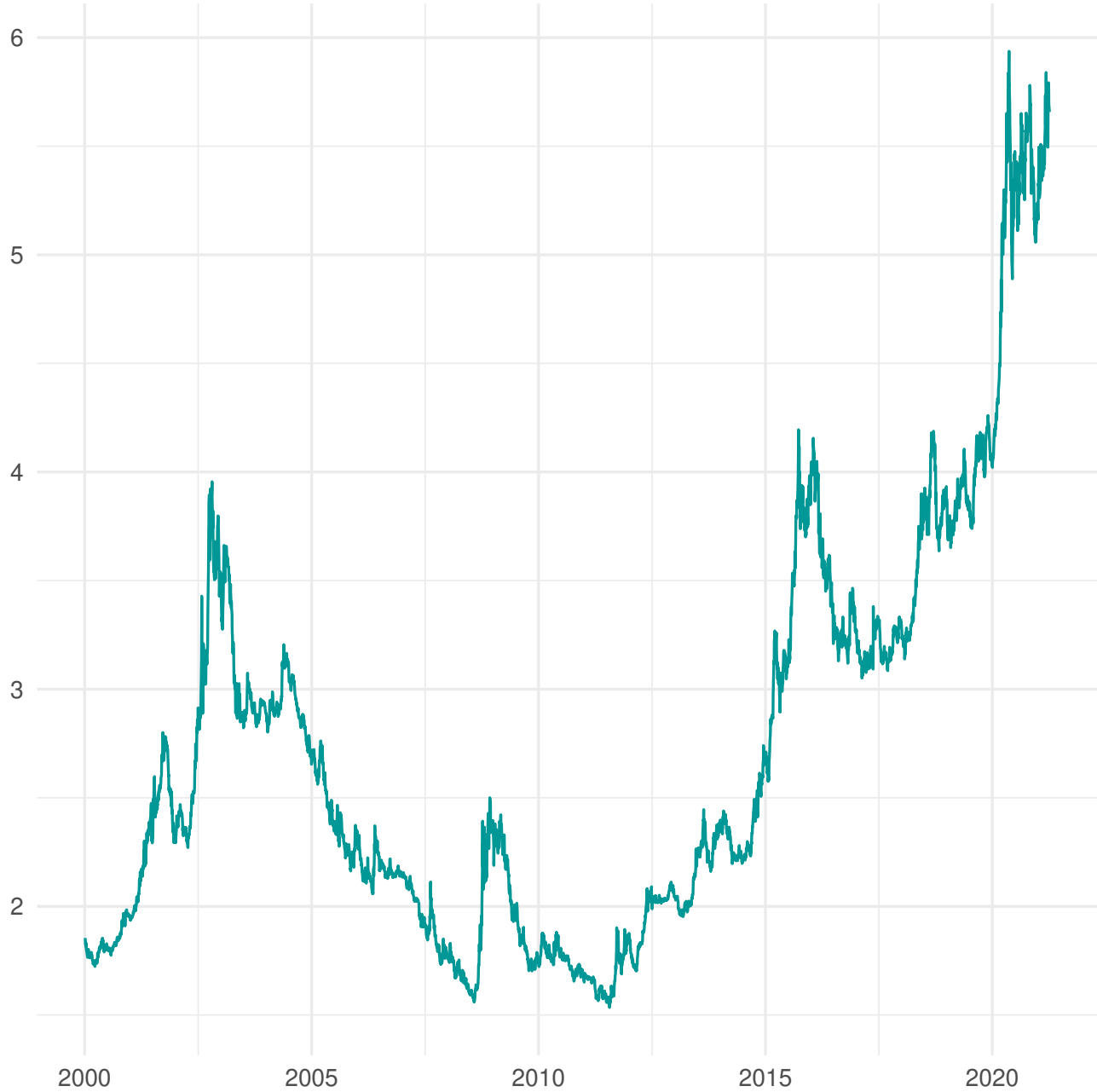
$$r_t = \log\left(\frac{P_t}{P_{t-1}}\right) = \log(1 + R_t).$$


Daily Closing Prices of Major European Stock Indices, 1991–1998
EuStockMarkets package:datasets
8000
6000
4000
2000
1993
1995
1997
1999
Series
DAX
SMI
CAC
FTSE

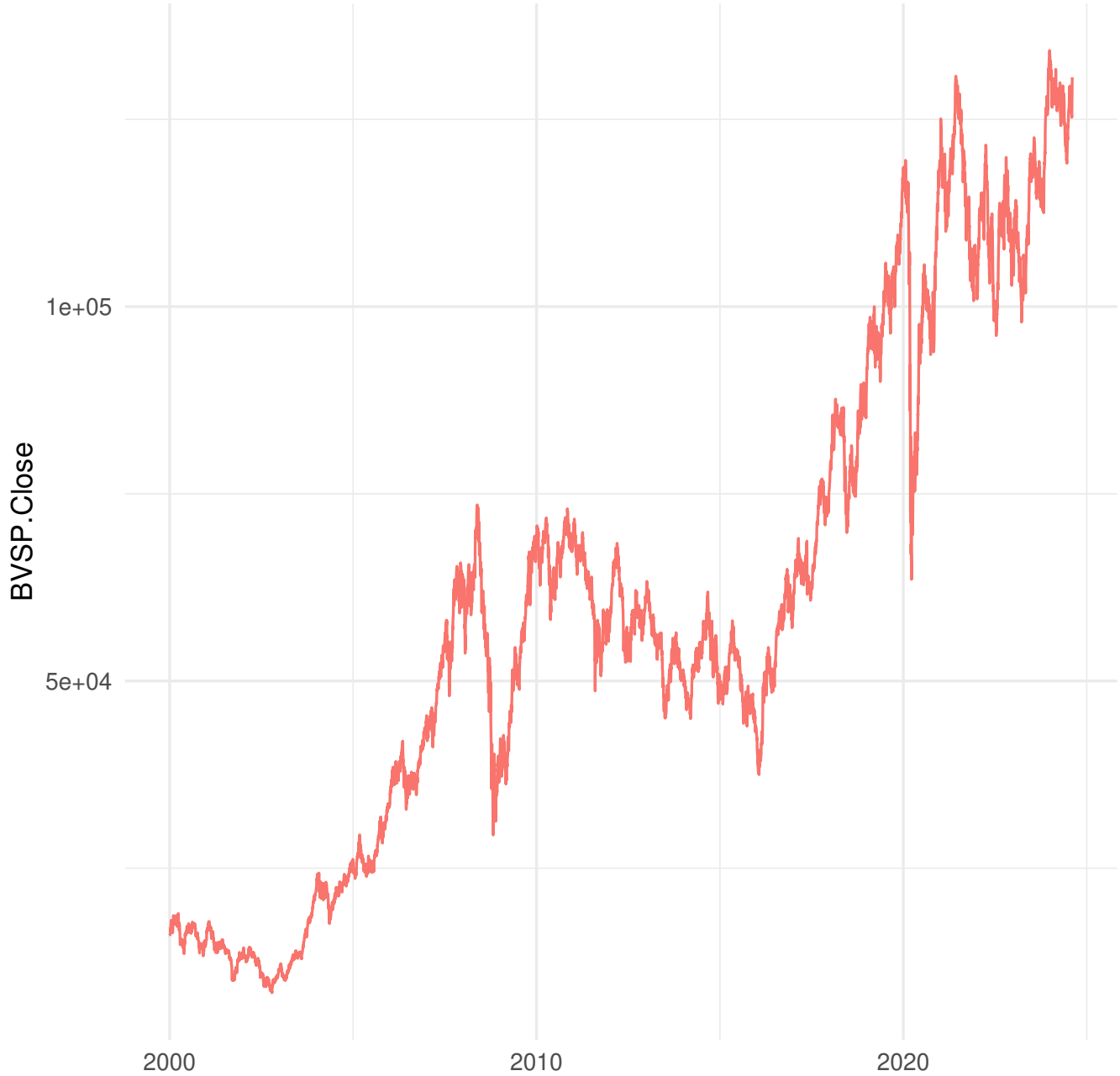

S&P 500 and Shanghai market indexes
index
5000
4000
3000
2000
1000
Shanghai
SP500
2010−01−01
2015−01−01
2020−01−01
2025−01−

Dow Jones Industrial Average (DJI) closing adjusted daily data (Yahoo Finance) from 05/01/2000 to 09/08/2024.

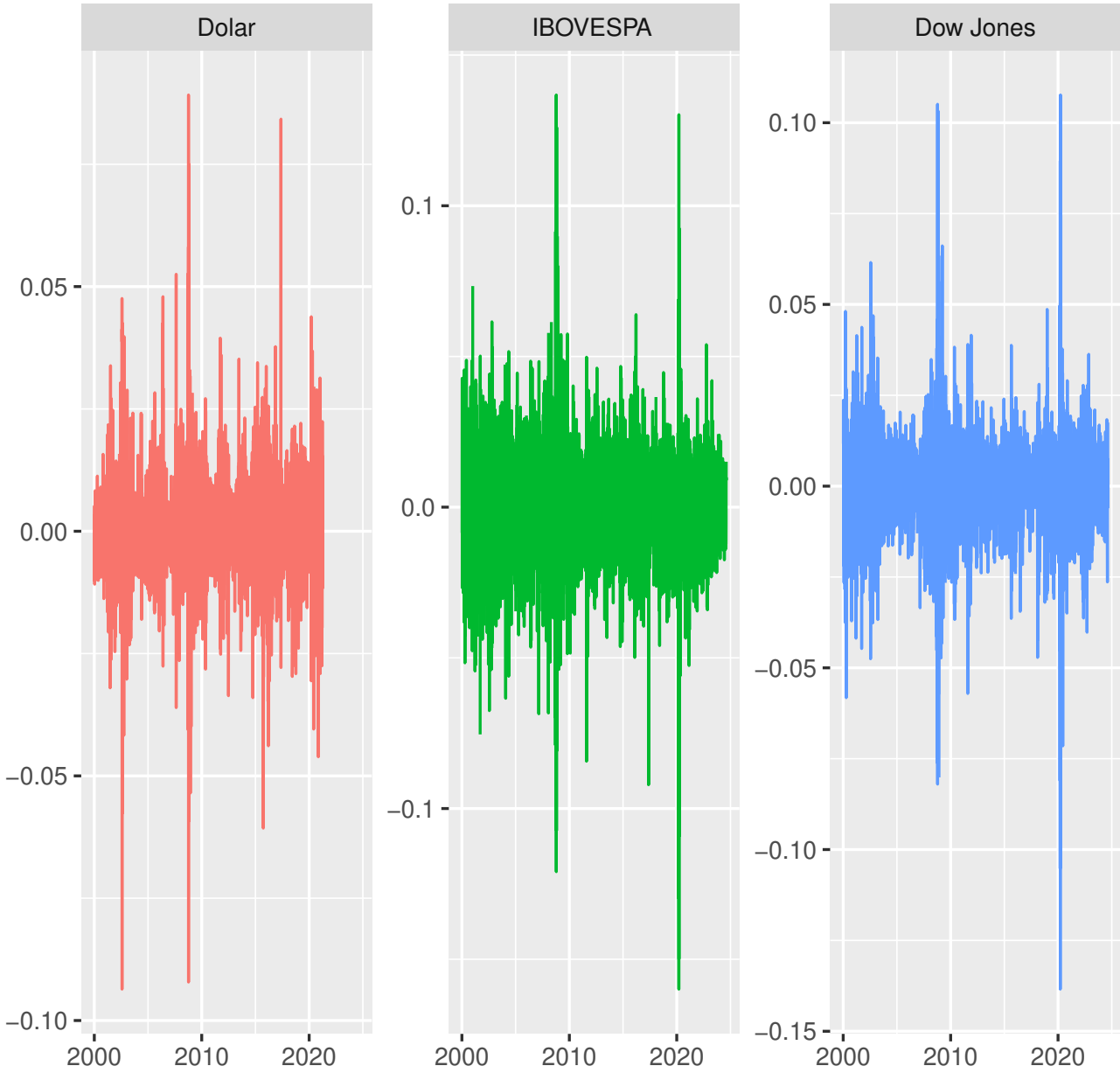
USD/BRL exchange rates from 05/01/2000 to 05/04/2021.



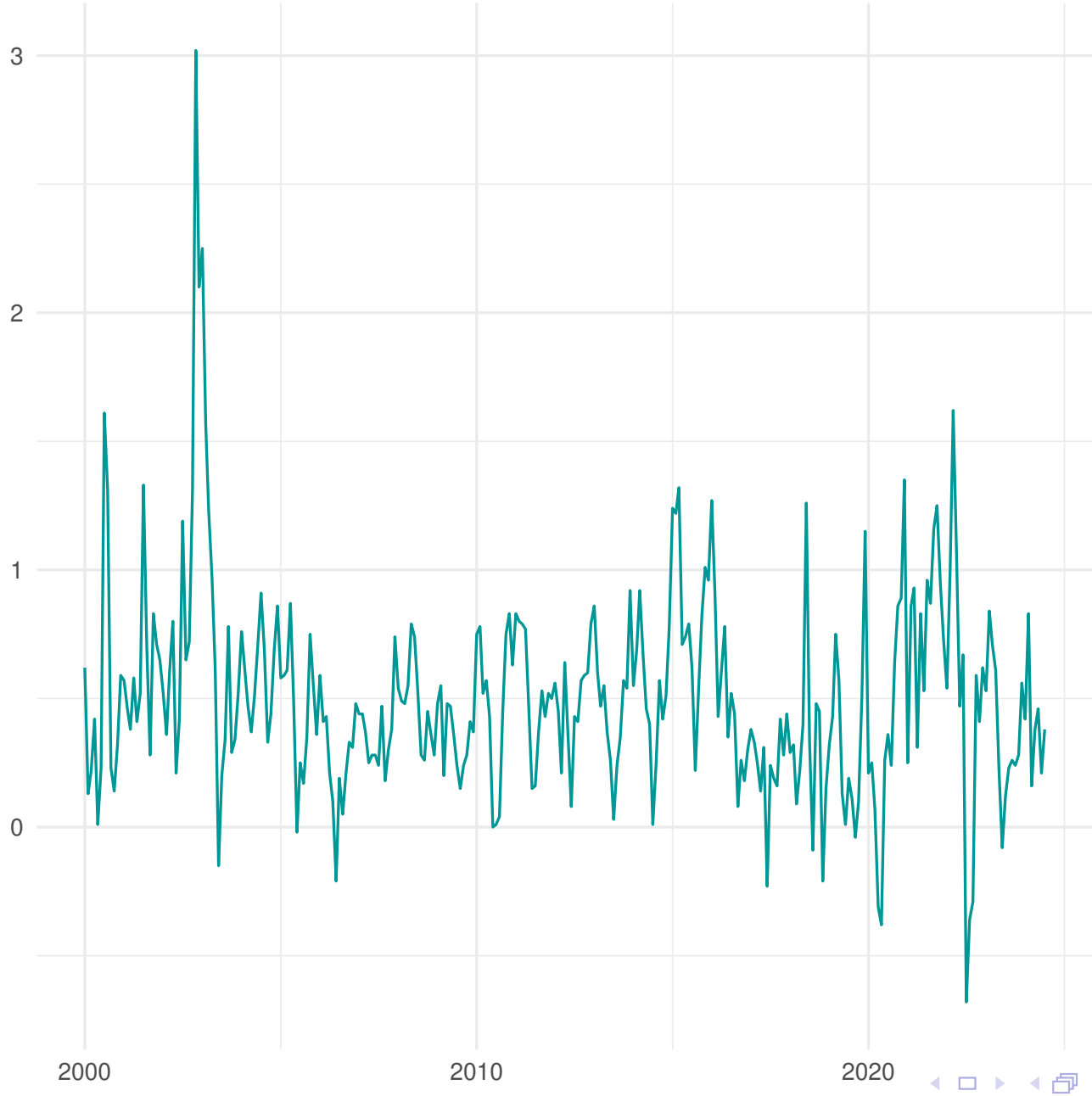
BOVESPA Index daily (closing) from 05/01/2000 to 09/08/2024.



Daily returns USD/BRL, IBOVESPA and Dow Jones



Monthly Extended National Consumer Price Index (IPCA) from jan 2000 to jul 2024.



Some Stylized Facts

- Returns tend to show heavier tails with a higher pick around the mean.
- Returns tend to show different variabilities along time.
- Returns variability tend to appear in clusters. Larger returns tend to be followed by larger returns. Likewise for small returns.
- Returns variability tend to increase more after a price decrease (negative return) than after a price increase of the same magnitude (leverage effect).

Read Chapter 1 of Prado & West