Air passengers 1949-1960

22 marks

Airline passengers.
Here we will examine the monthly total of the number of air passengers in the US from 1949 to 1960.

The data are available as the data set `AirPassengers` from the `datasets` package in the standard `R` distribution.

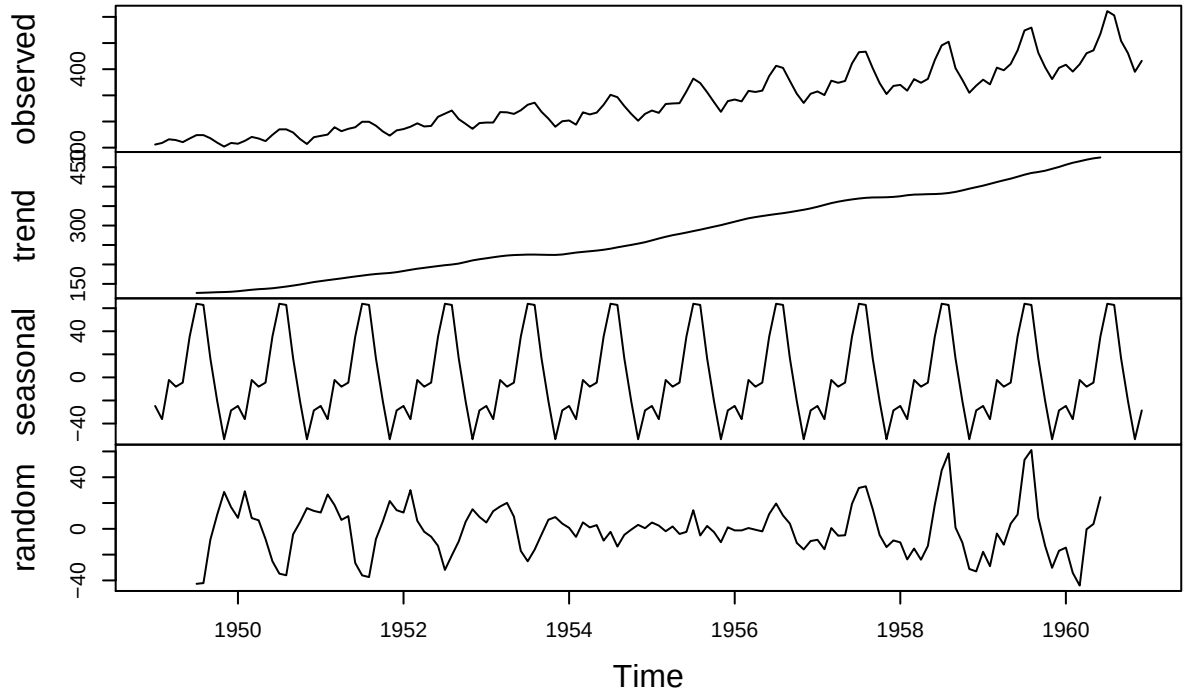
```
AirPassengers
```

```
##      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
```

a. **(7 marks)** Use the function `decompose()` on this data and plot the results.

   i. *(2 marks)* Show your plot.

```
plot(decompose(AirPassengers))
```



ii. *(1 mark)* Describe the trend.
There is an rising trend with the time increases.

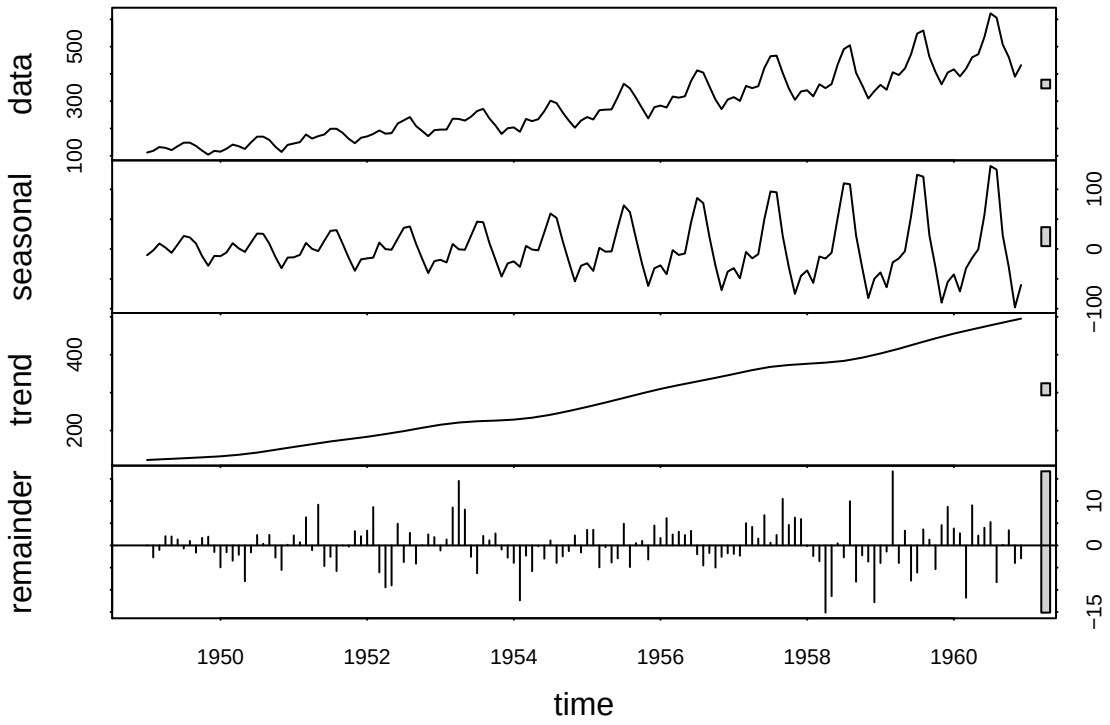
iii. *(2 marks)* Describe the seasonal pattern.
The seasonal is in a repeating cycle across years. Graphs are indentical every year.

iv. *(2 marks)* What do you conclude from the residuals?
We don't see a clear pattern in residuals. However, the plot between 1954 and 1958 is different from other years.

b. **(9 marks)** Use the function `stl()` on this data with seasonality loess span `s.window = 7` and the local polynomial for the seasonal loess being a line.

i. *(2 marks)* Show your plot.

```
plot(stl(AirPassengers, s.window=7, s.degree=1))
```



ii. *(2 marks)* Describe the trend. Is it significant? Why or why not?

It's still an rising trend. However, it is significant because the variation is really large as the size of the range rectangle is smaller than the remainder's.

iii. *(3 marks)* Describe the seasonal pattern. Is it significant? Why or why not?

Seasonal variation gets larger and larger but still in certain cycles. It is significant because comparing to remainder, it also has small size of range rectangle.

iv. *(2 marks)* What do you conclude from the residuals?

We can't see a clear pattern in residuals. There are variations across years.

c. **(6 marks)** Compare the plots from parts (a) and (b).

i. *(2 marks)* What are the major differences in the plots?

Seasonal variation in (b) increases across years comparing to constant variation in (a).

ii. *(2 marks)* What characteristics of the two methods caused these differences?

It is because for (a), we are using simple average for months in all years. In (b), we are smoothing the value in a 7-year window and use the smooth values as estimate.

iii. *(2 marks)* Which fitted model do you prefer and **why**?

I prefer the second one because we are able to find a descriptive pattern in seasonal. Residuals in the second model is random, which is a desired result.