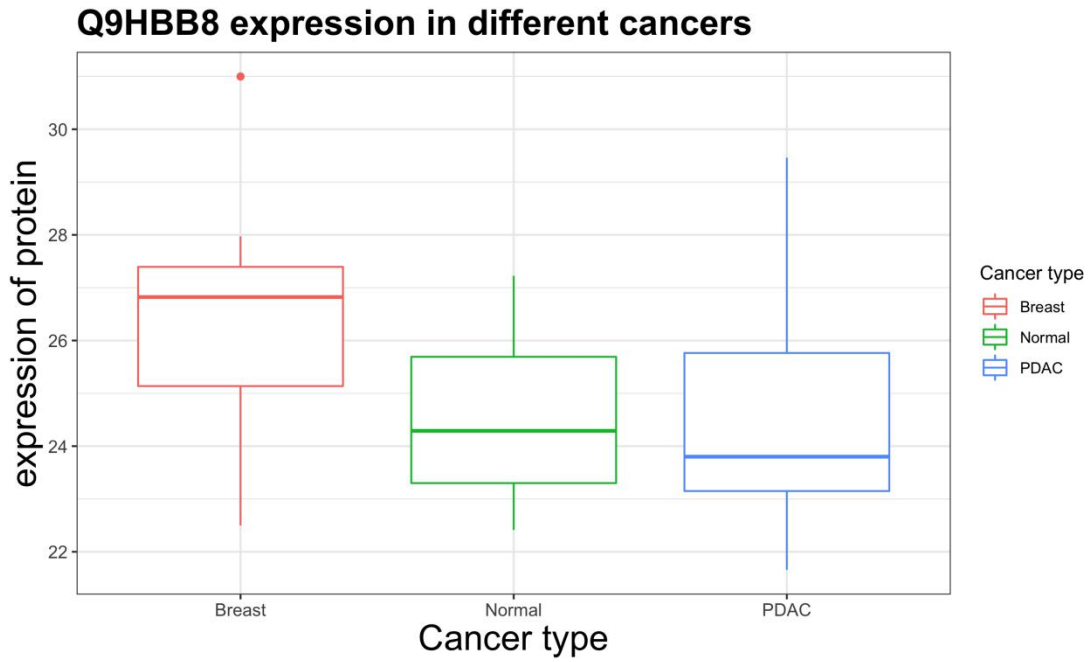
Assignment 04

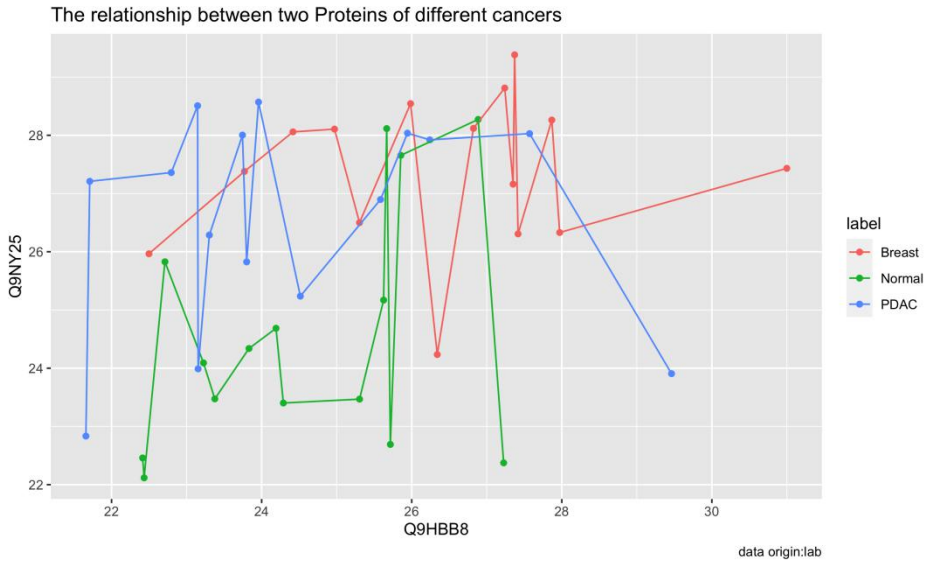
1. Plotting with ggplot2

[25 points] Using research data from your group, make 5 types of plots with the ggplot2 package:

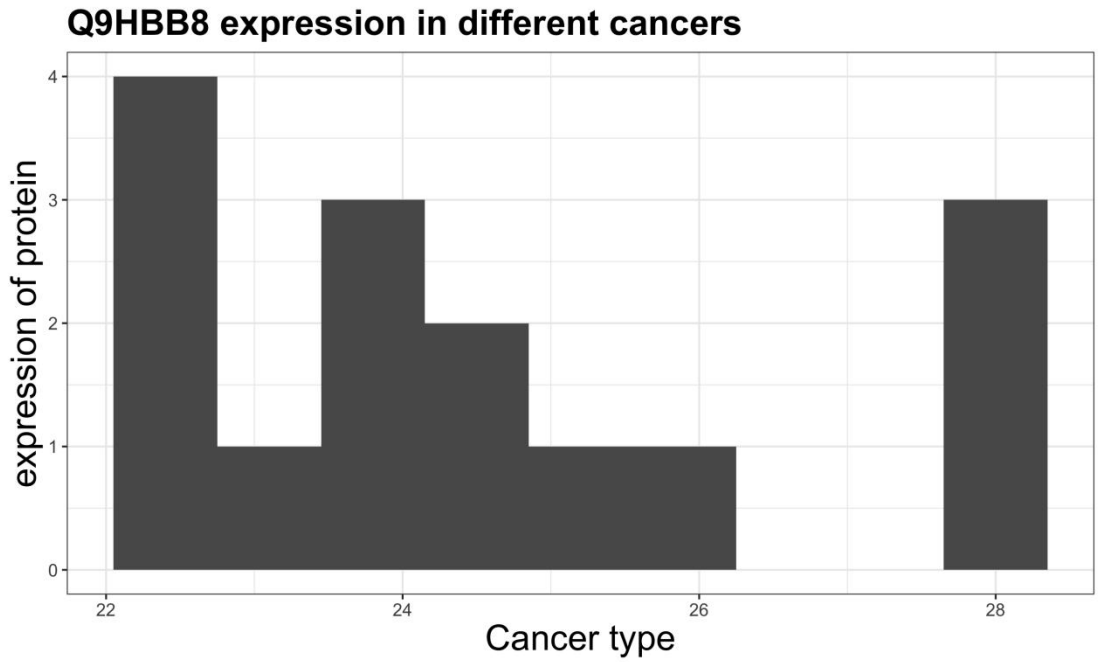
Boxplot



Time series



Histogram



Scatter plot

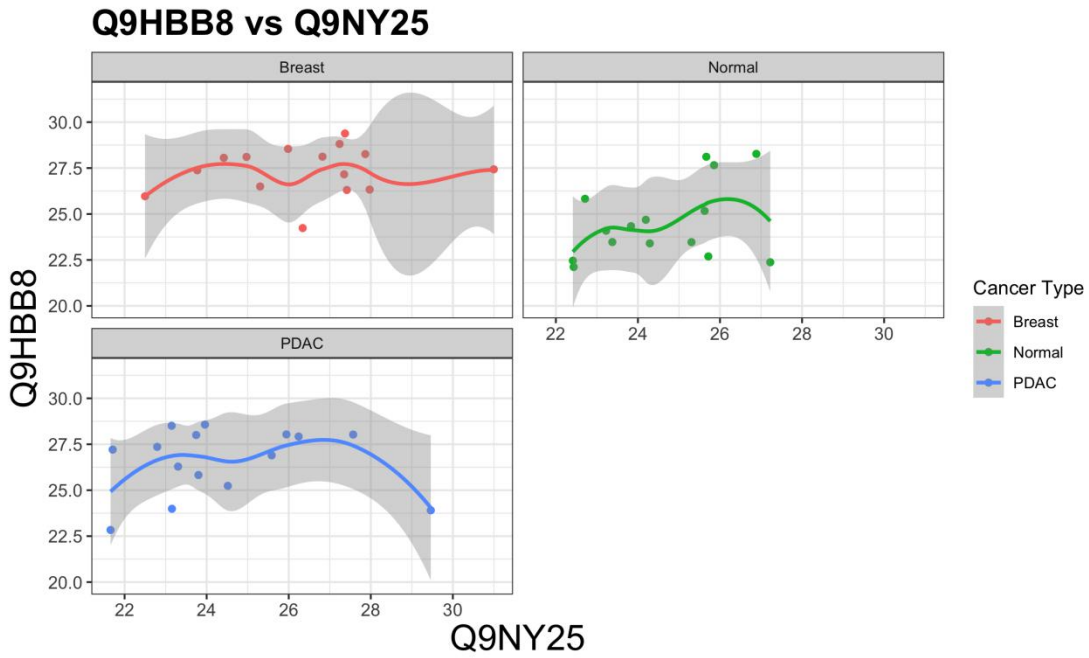
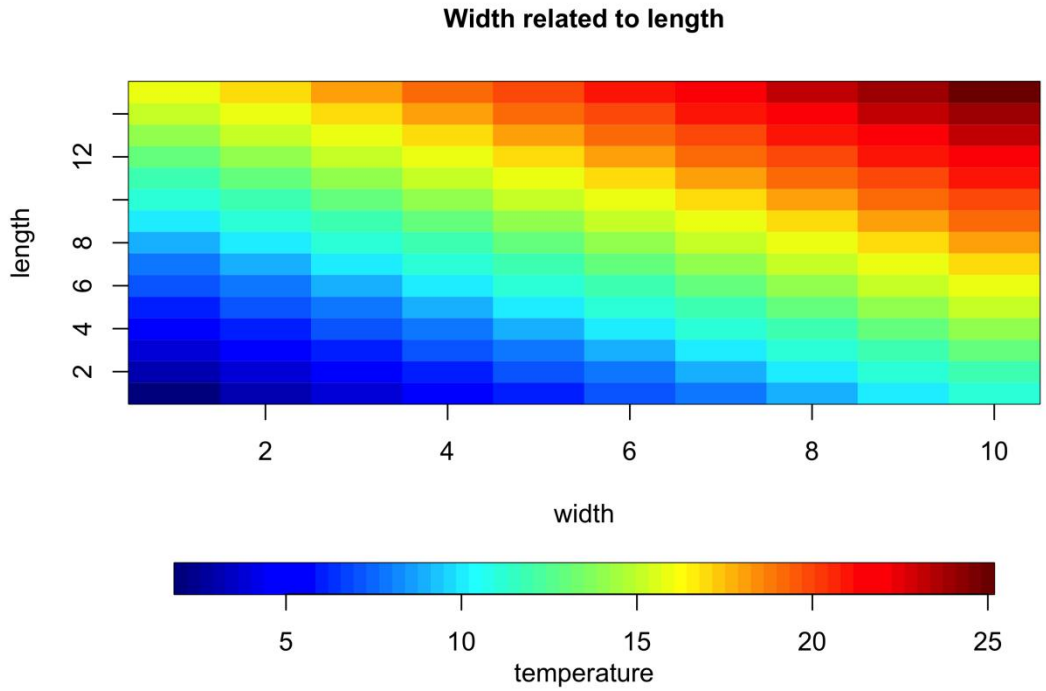


Image plot (you can use data set of interest for this one)

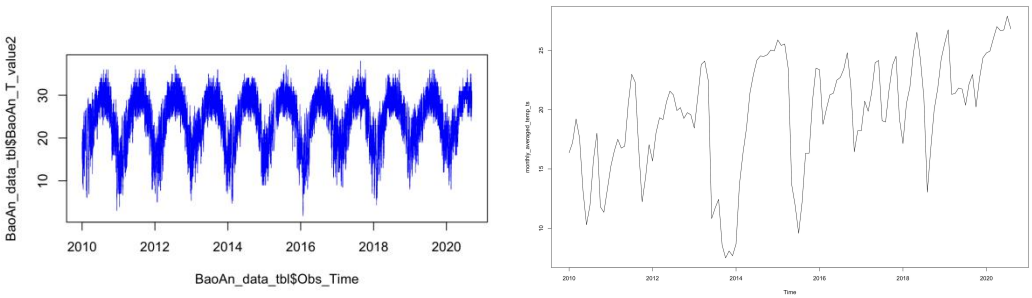


2. Analysis of the time series of monthly temperature

In this exercise, we will take another look at the hourly weather data measured at the BaoAn International Airport during the past 10 years.

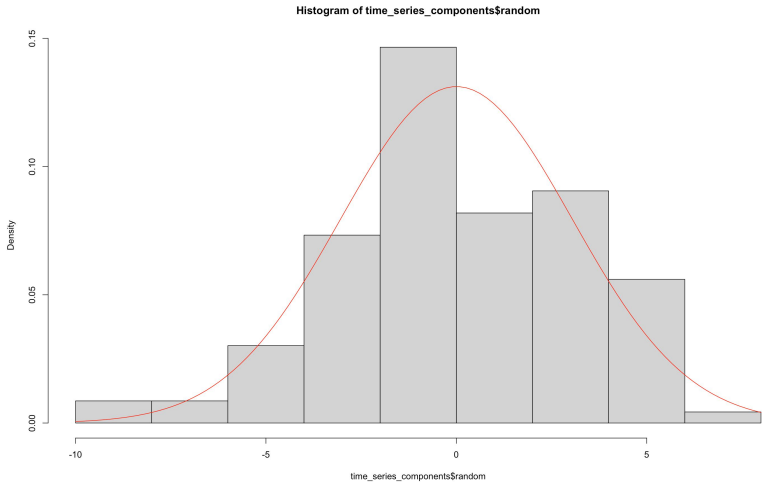
2.1 [5 points] Construct a time series of monthly-averaged temperature from 2010 Jan. to 2020 Aug.

Answer: I can get monthly temperature as same as the submitted home work with package “dplyr” . After we get clean data, we can use “ts()” function to build time series model from 2010 Jan. to 2020 Aug.



2.2 [5 points] Decompose the time series into trend, seasonality, and error parts. Check whether the error part follows a white noise distribution.

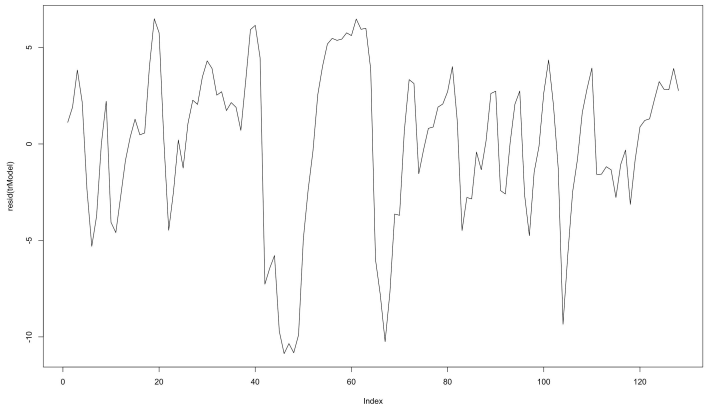
Answer: I use decompress() to get the main component of time series data.



2.3 [10 points] Fit an ARIMA(p,d,q) model to the time series. Describe the fitting process in details in your report.

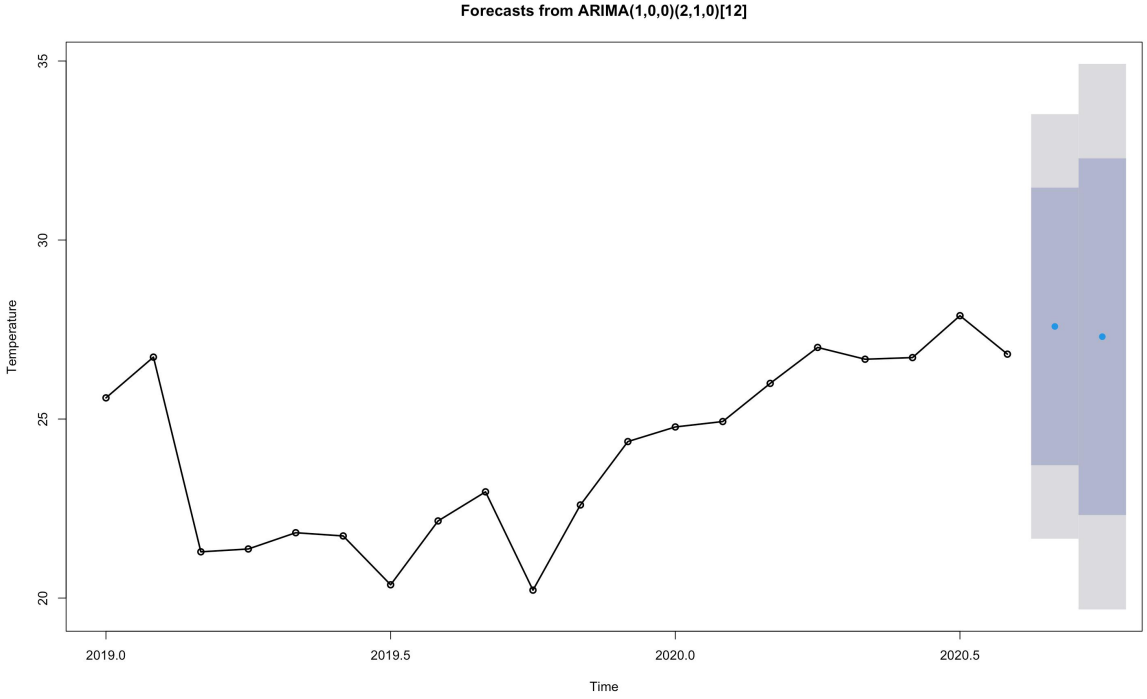
Answer: I construct a linear model with “lm” function, and then check “acf and pacf” of this model, after that, I conducted automated forecasting with using an ARIMA model, final, I use “arima” to help me find a better model, and I find “order=c(2,1,0)” has better predict ability, so I decide use this model in the following analysis.

Best model: ARIMA(1,1,2)
arima1: aic = 618.65
arima2: aic = 602.18



2.4 [5 points] Predict monthly-averaged temperatures in 2020 Sep. and Oct. with the ARIMA model from 2.3. The predictions will be evaluated against actual observations in those two months.

Answer: I use forest function to forest the following two months with arima2 model


Forecasts from ARIMA(1,0,0)(2,1,0)[12]
Temperature
35
30
25
20
2019.0
2019.5
2020.0
2020.5
Time