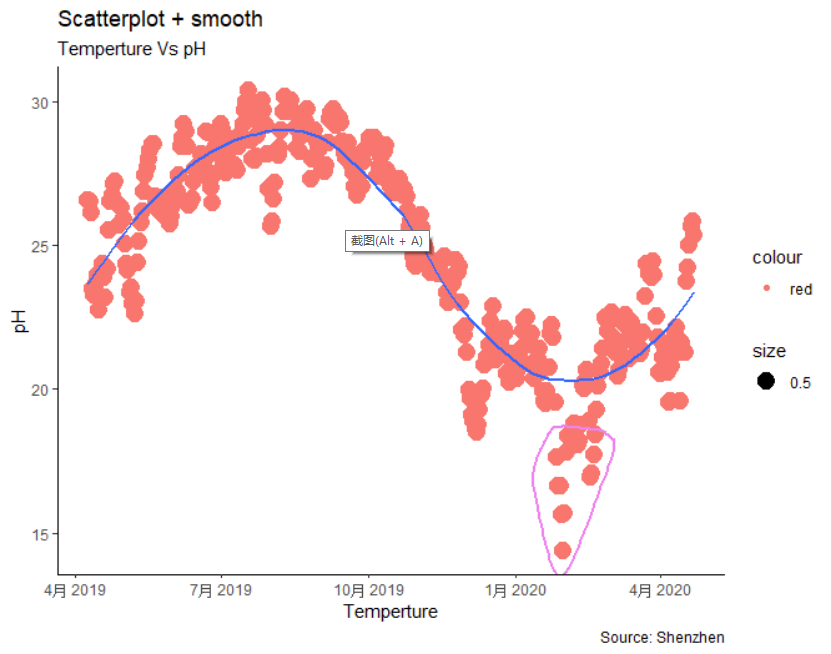
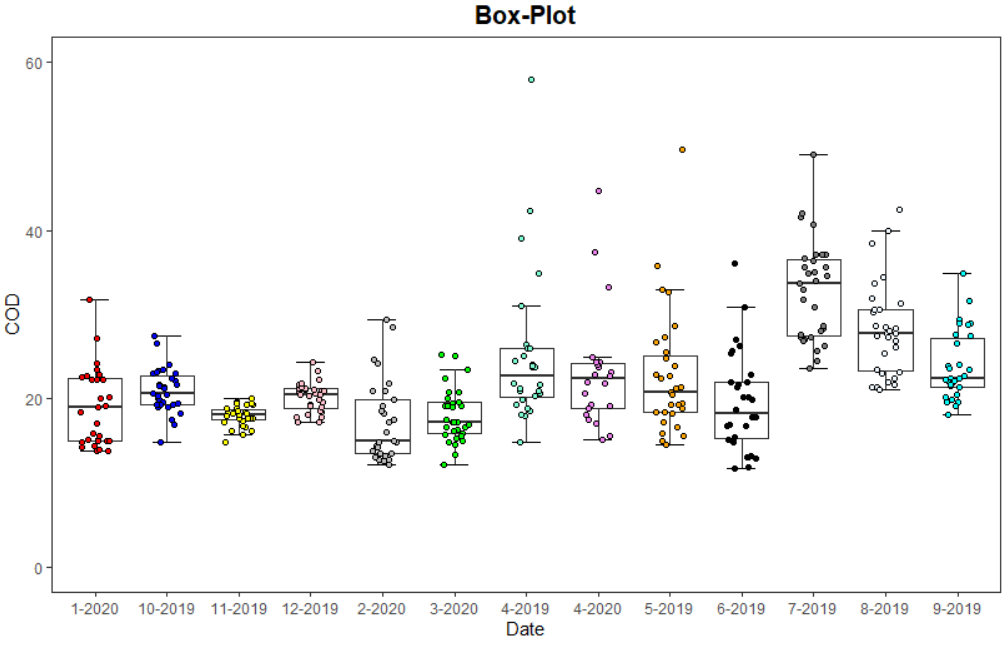
**Problem 1**

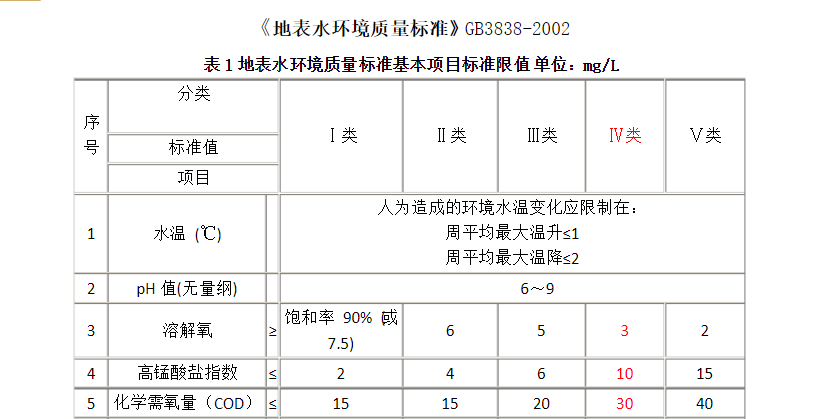
散点图



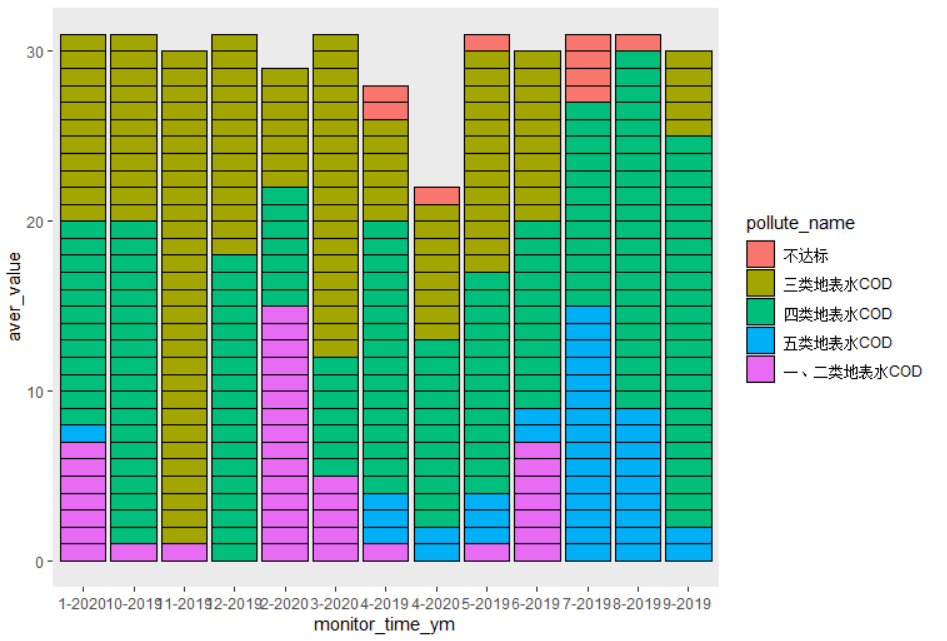
箱型图



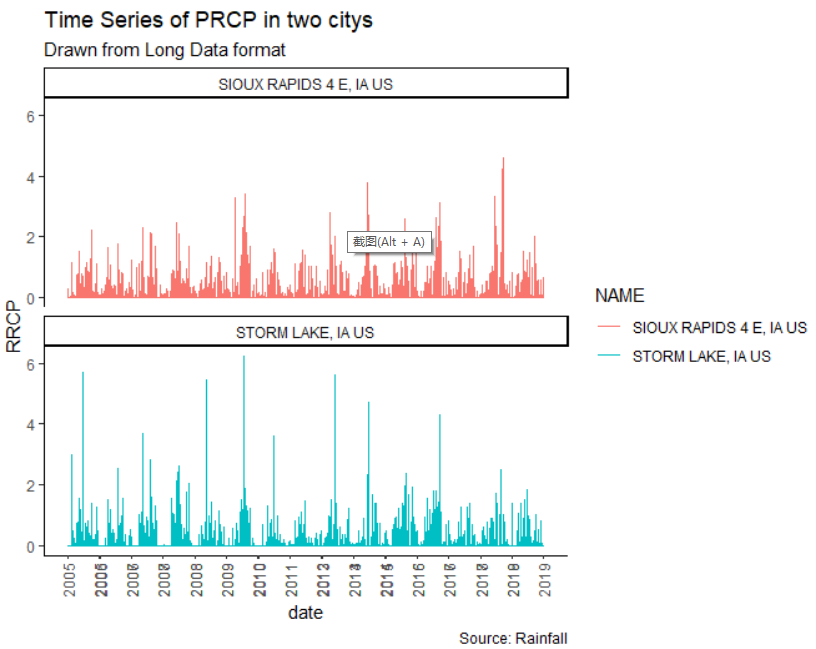
直方图



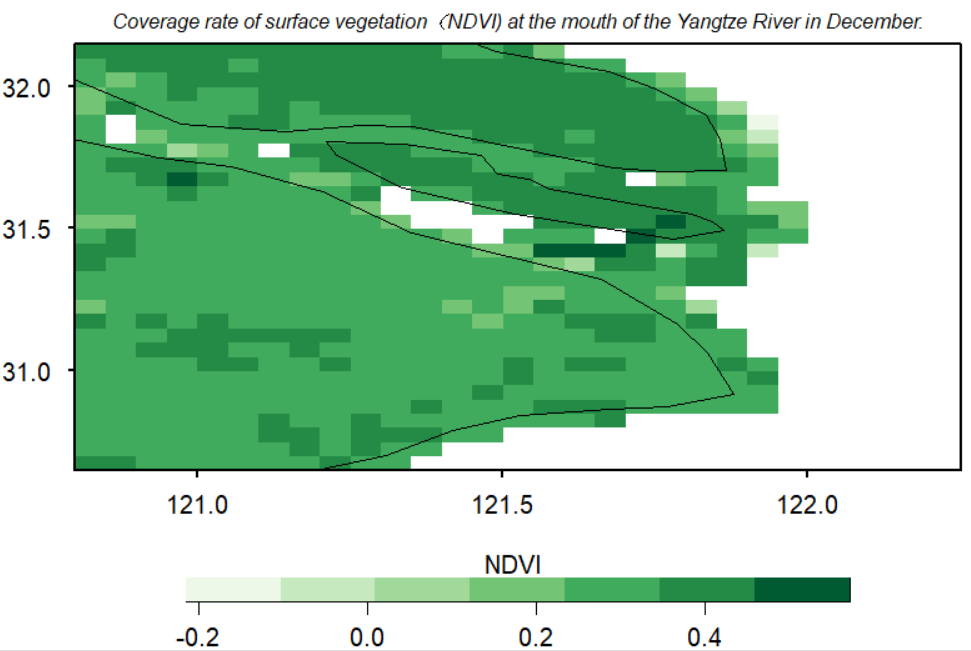
根据地表水环境质量标准对深圳市某河的COD指标进行评价。做每月的堆叠直方图。



时间序列图



图像图



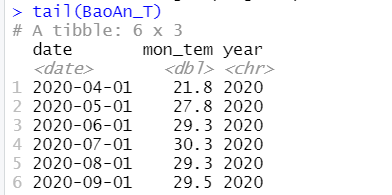
**Problem 2**

#2.1

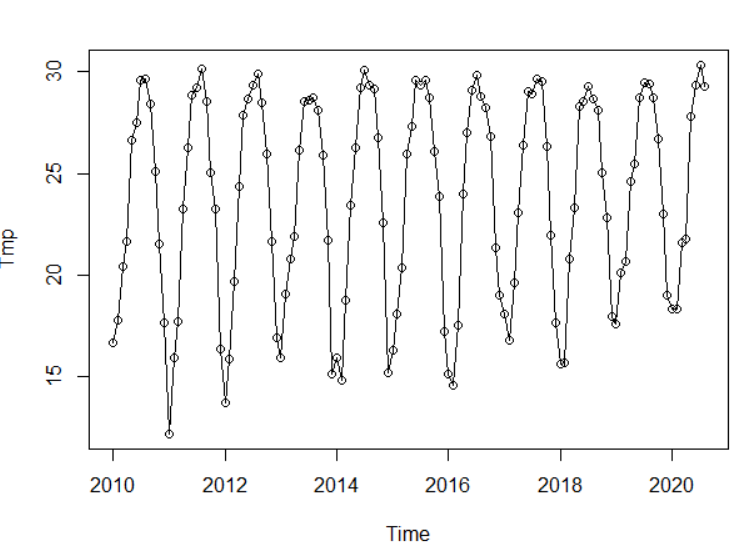
首先载入程辑包，读取数据，并转化为tibble型。

然后对温度数据进行筛查，剔除掉温度为9999的异常值，并且温度标识不为1的温度数据也要剔除。

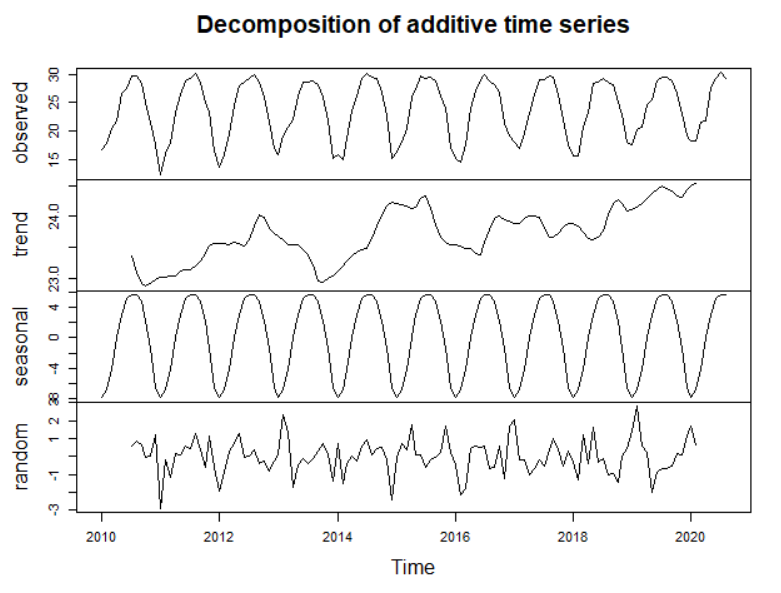
接着将日期数据转化为日期型，要注意日期型数据由年月日组成，为方便求月平均温度，我将日统一定为一号。同时，截取温度数据。因为温度数据没有小数，经观察数据得，需要将截取的数据转化为数值型，并除以10.本题所提供的原始数据并没有10月的数据，我从“胡兆平”处获得了他下载的完整的数据。

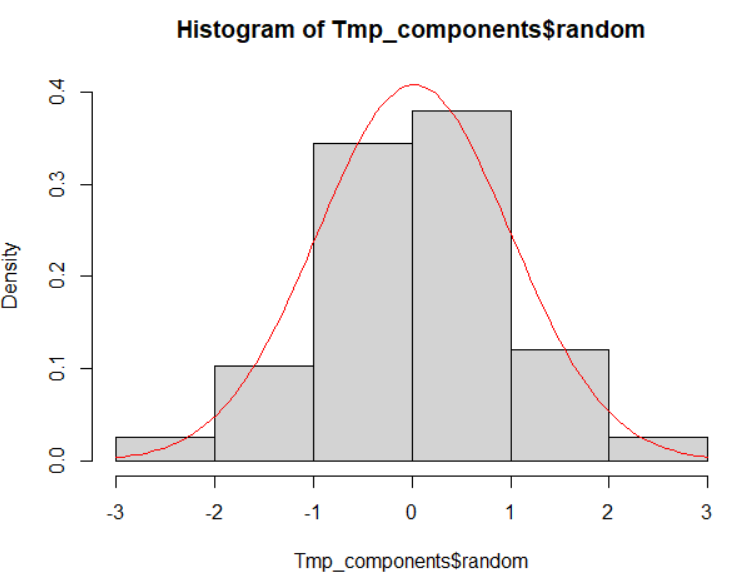


构造一个从2010年1月到2020年8月的月平均气温的时间序列。如图



#2.2

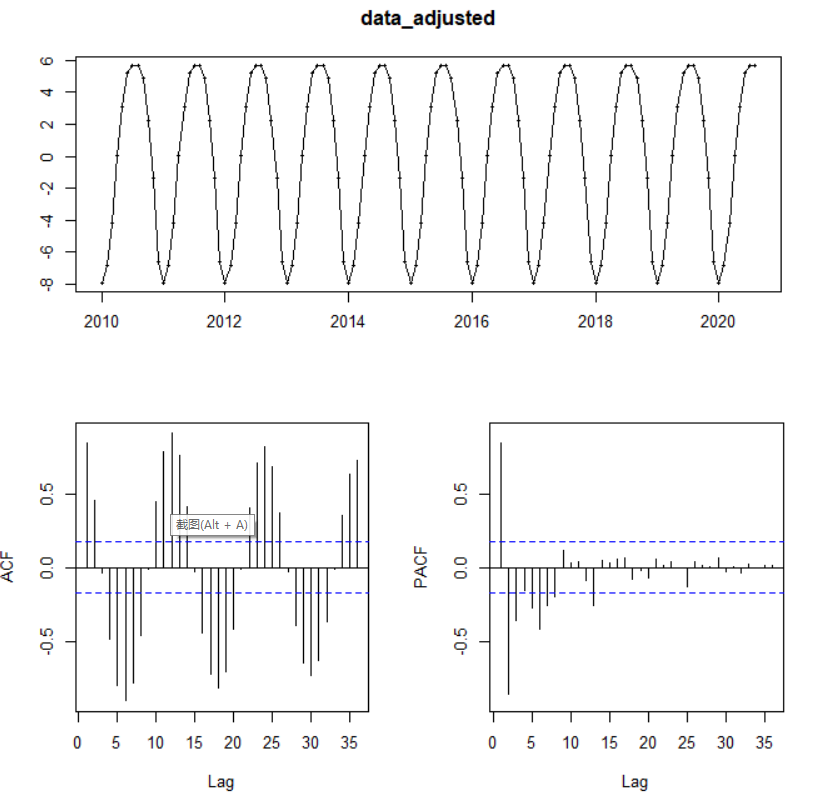




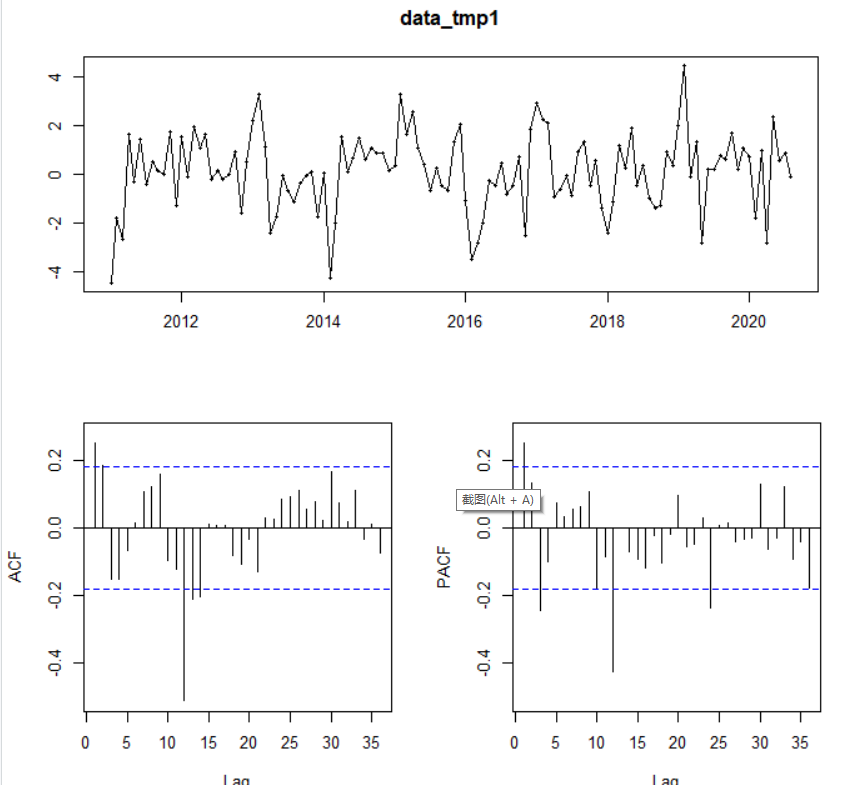
将时间序列分解为趋势，季节性和误差部分。通过检查错误分量的分布以确定是否遵循白噪声分布。如图所示，分布为高斯白噪声。

#2.3

首先我们注意到组分中的季节部分是呈现周期性的，这里我们忽略季节性的影响，对模型的季节性因素做1阶12步差分。差分前：



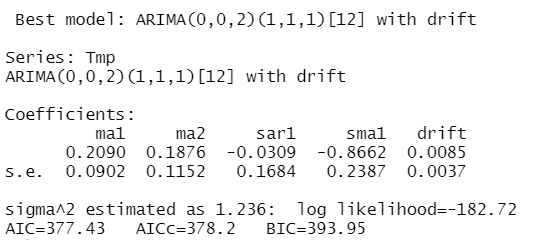
差分后：



模型基本稳定

平稳后就是确定ARIMA模型的阶数了

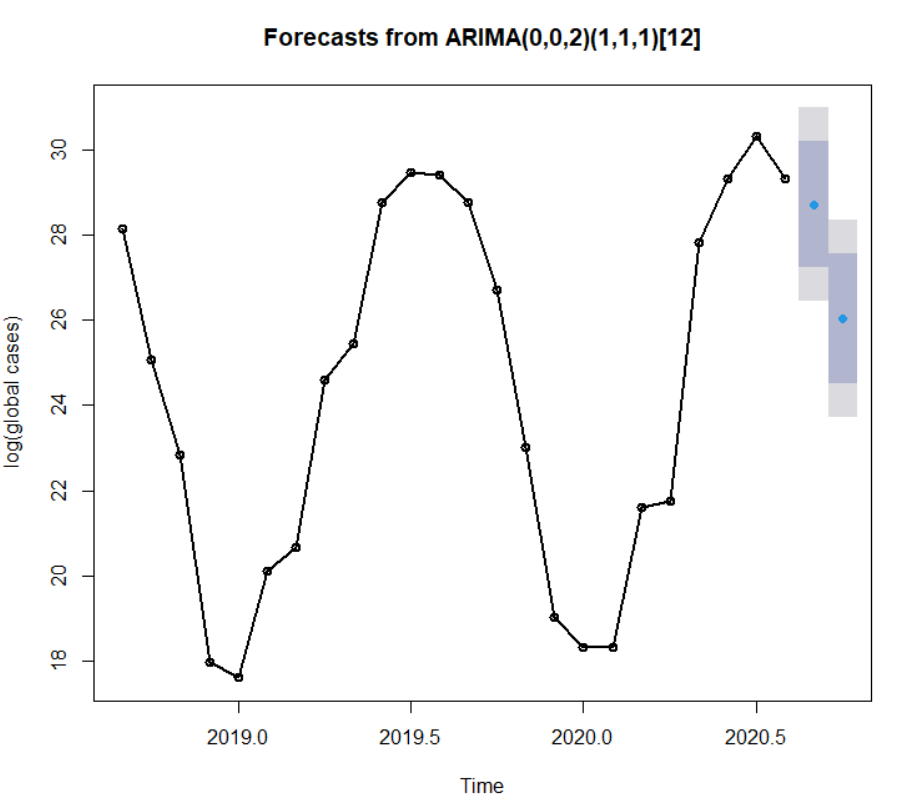
这里用forecast包的auto.arima()自动识别最优的阶数



#2.4

预测9和10月的月平均气温，并将其与实际值绘制在一起。





再从别的数据文件读入10月数据，并进行简单处理，所以与九月月平均气温的实际值比较，相对误差仅为0.03%，与10月数据相比，相对误差-0.09%，模型预测结果较好。