PS4 Report

吴文浩 12032567

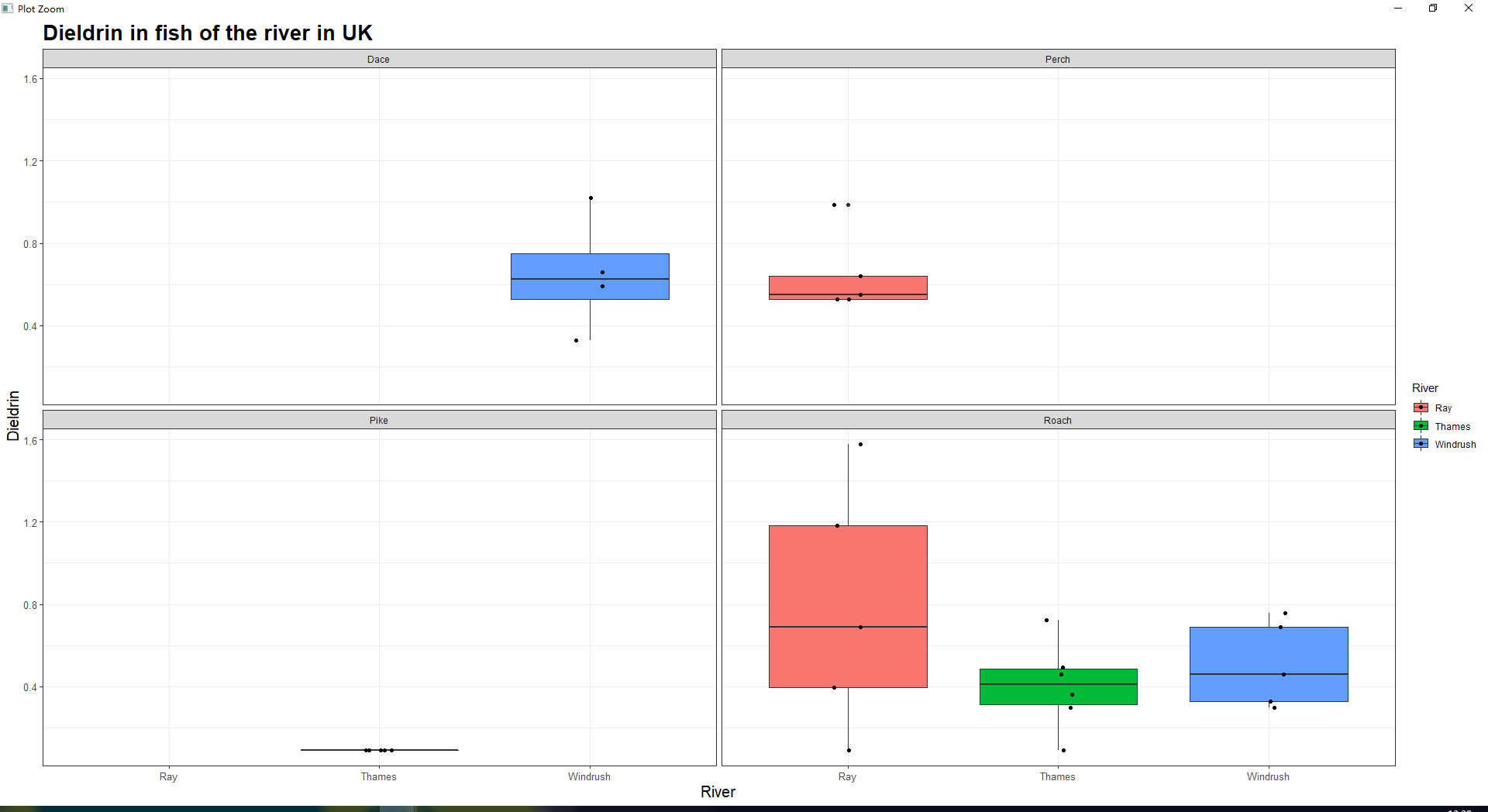
1. PS4\_1

**注：**由于所研究方向为无人机设计，没有观测数据，故采用R语言自带数据集以及NADA数据集，已与老师提前报备

**1.1：箱型图：**选用NADA包中的数据集Thames，Thames数据集反应了化合物Dieldrin在英国各条河流中，三种鱼类体内的含量

**箱型图美化包括：**标题，坐标轴，根据河流填充箱型图色彩，根据鱼类划分四个数据框，增加数据点在箱型图上的排布

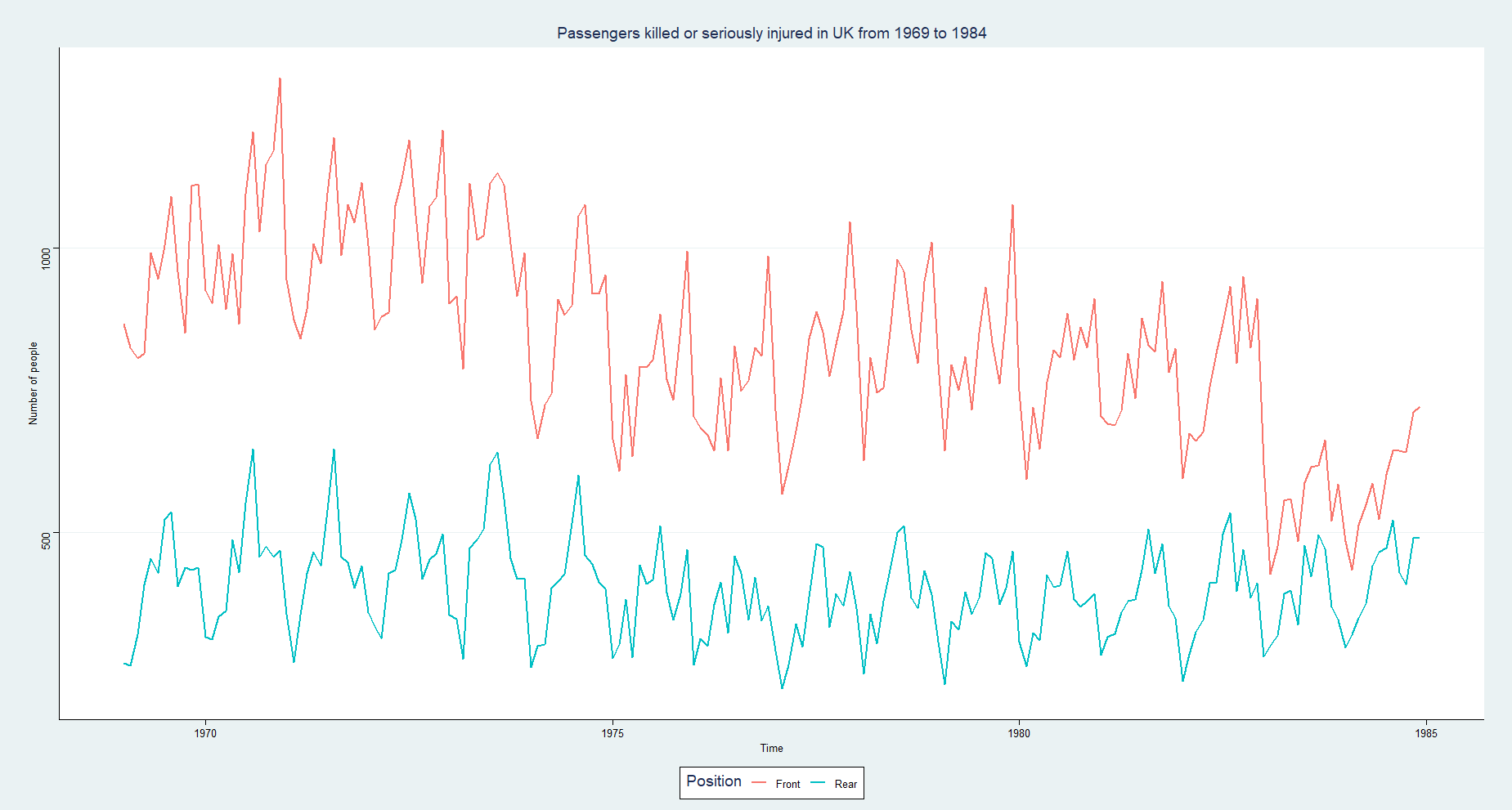
**结果截图：**



**1.2：时间序列：**采用R自带数据集中的多参量时间序列数据集Seatbelts，记录了1969-1984年期间英国的交通事故情况，这里选用其前排乘客及后排乘客伤亡情况做时间序列对比图

**时间序列图美化包括：**标题，坐标轴，线型修改，更改ggplot主题风格，根据座位位置绘制不同颜色序列线

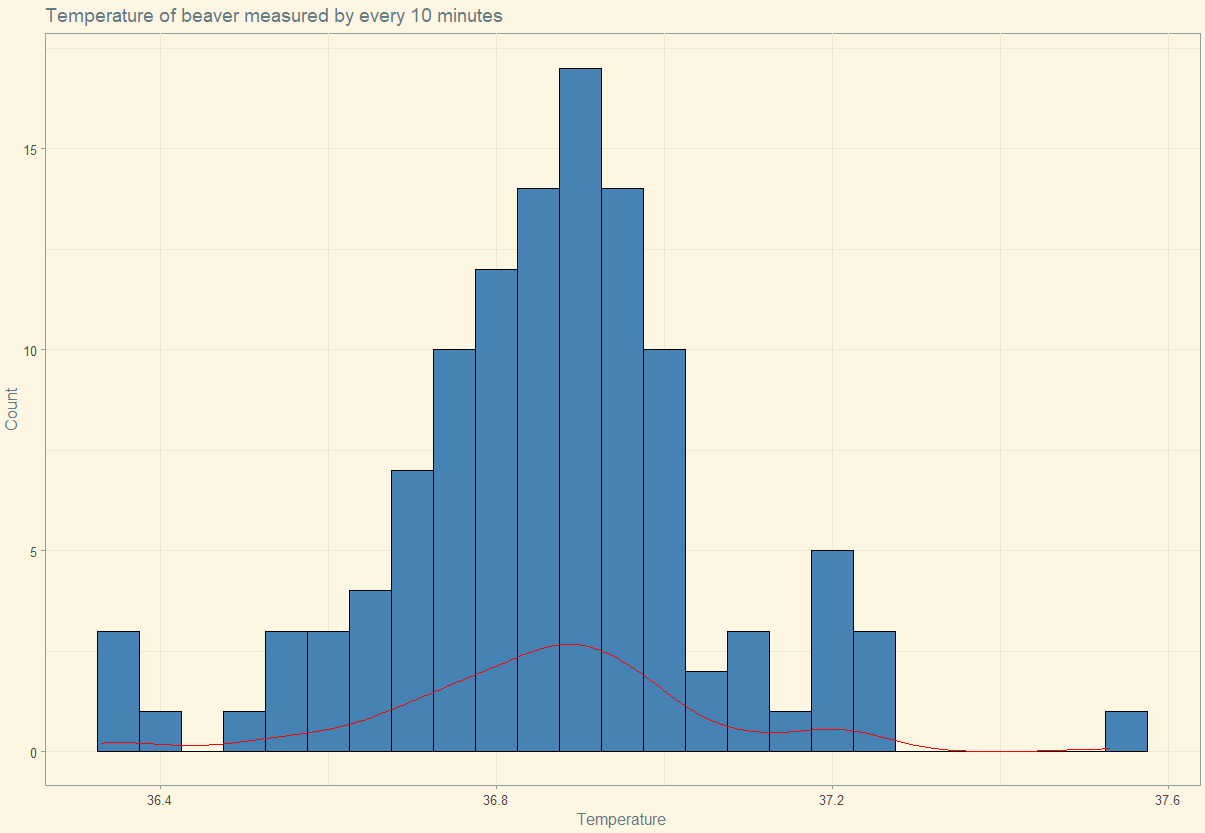
**结果截图：**



**1.3：直方图：**采用R自带数据集中beaver1，记录了某只海狸每10分钟的体温数据

**直方图美化包括：**标题，坐标轴，增加PDF曲线，更改ggplot主题风格，修改直方图步长

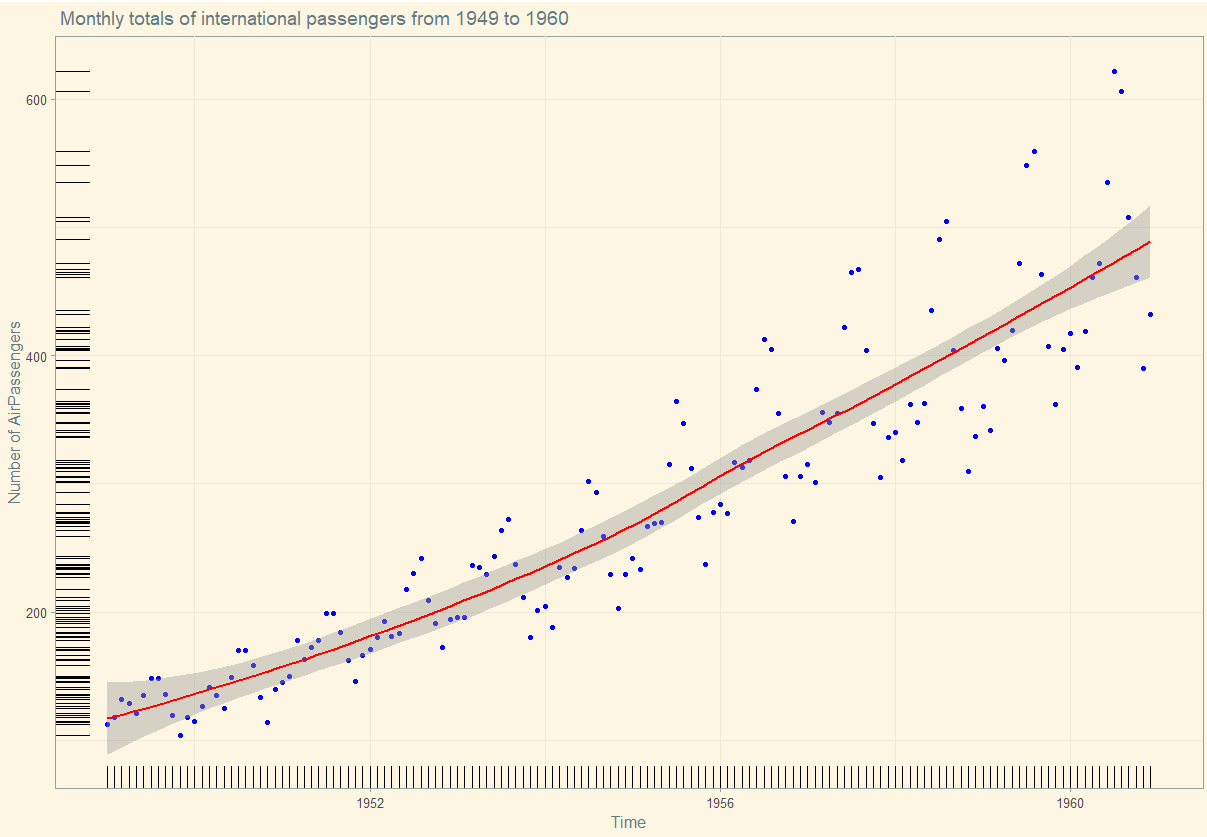
**结果截图：**



**1.4：散点图：**采用R自带数据集中AirPassengers时间序列，记录了1949-1960年期间的国际航班乘客数量

**散点图美化包括：**标题，坐标轴，增加边际地毯刻度尺，更改ggplot主题风格，增加回归曲线，修改坐标点颜色

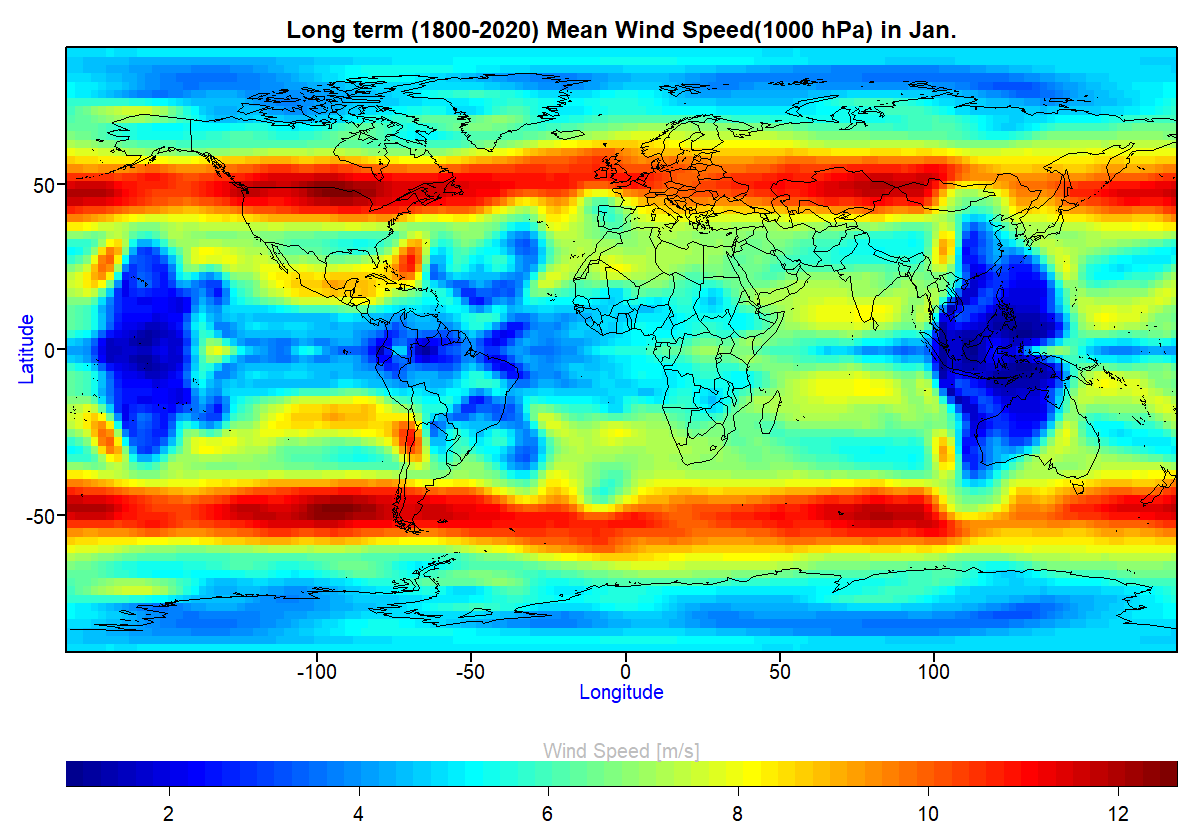
**结果截图：**



**1.5：Image plot：**采用noaa上的1800-2020每月平均风速netcdf数据进行绘制，由于该netcdf上的经度范围是0-360，需要修改切换会-180—180的范围进行显示（由于map函数导入的地图其经度范围是-180—180）

**Image plot美化包括：**标题，坐标轴，坐标轴名称颜色，legend名称颜色，字体大小

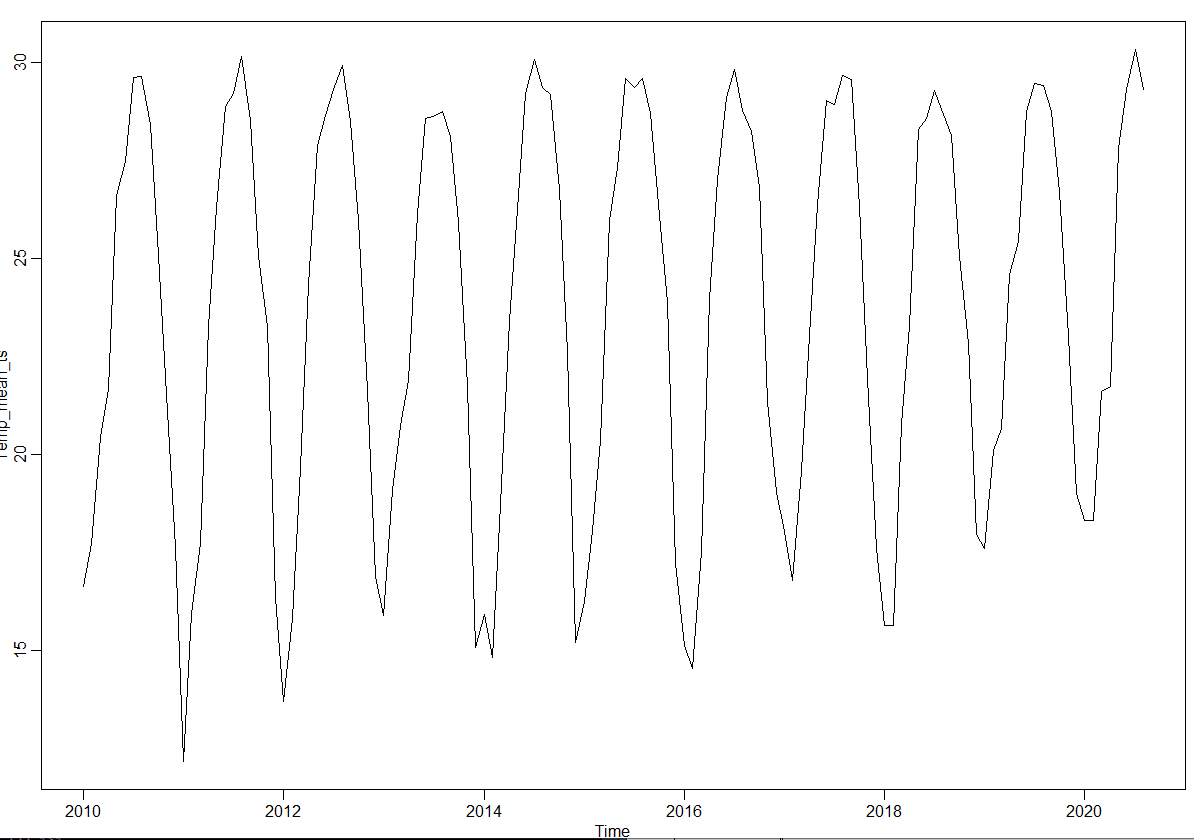
**结果截图：**



1. PS4\_2

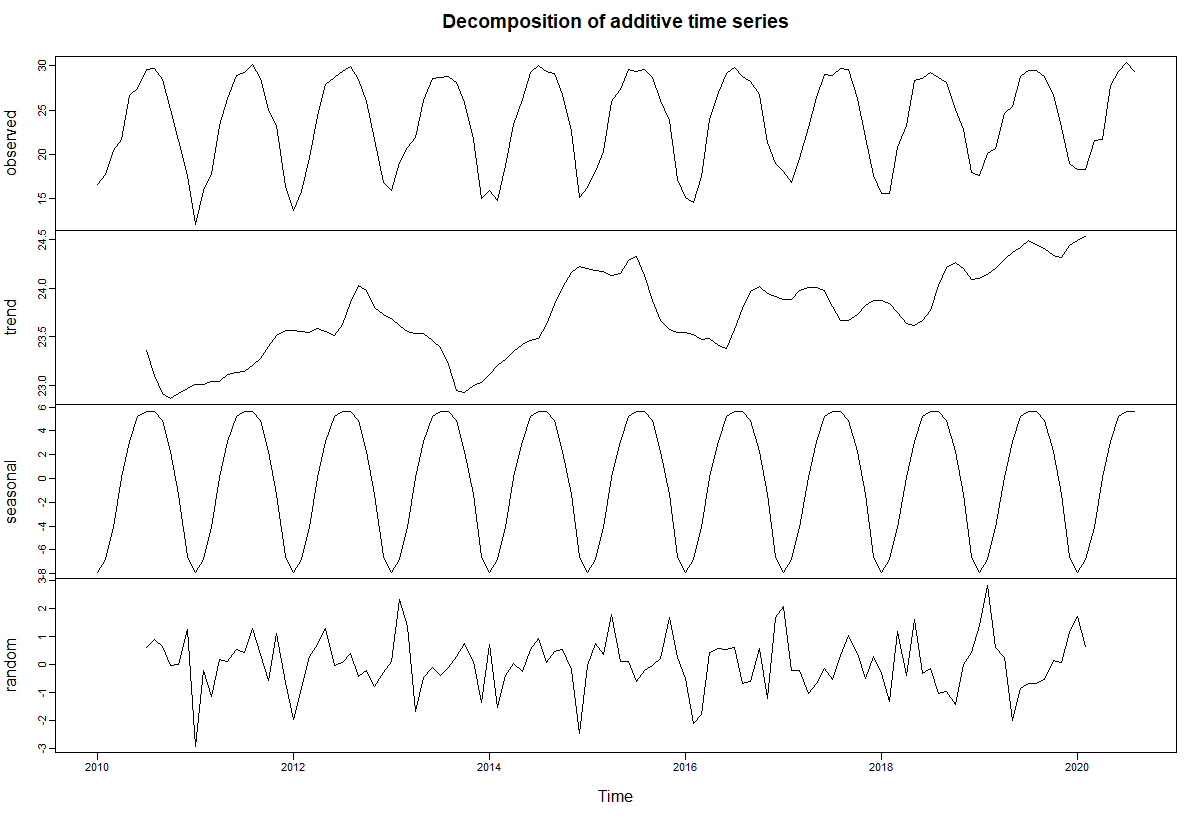
**2.1：**读取数据并转换为tibble格式，将其中无效数据进行过滤，并将日期转换年月的组合，进行过滤只保留2010年10月到2020年8月期间的数据，整合求解每月平均气温，并转化为实践序列

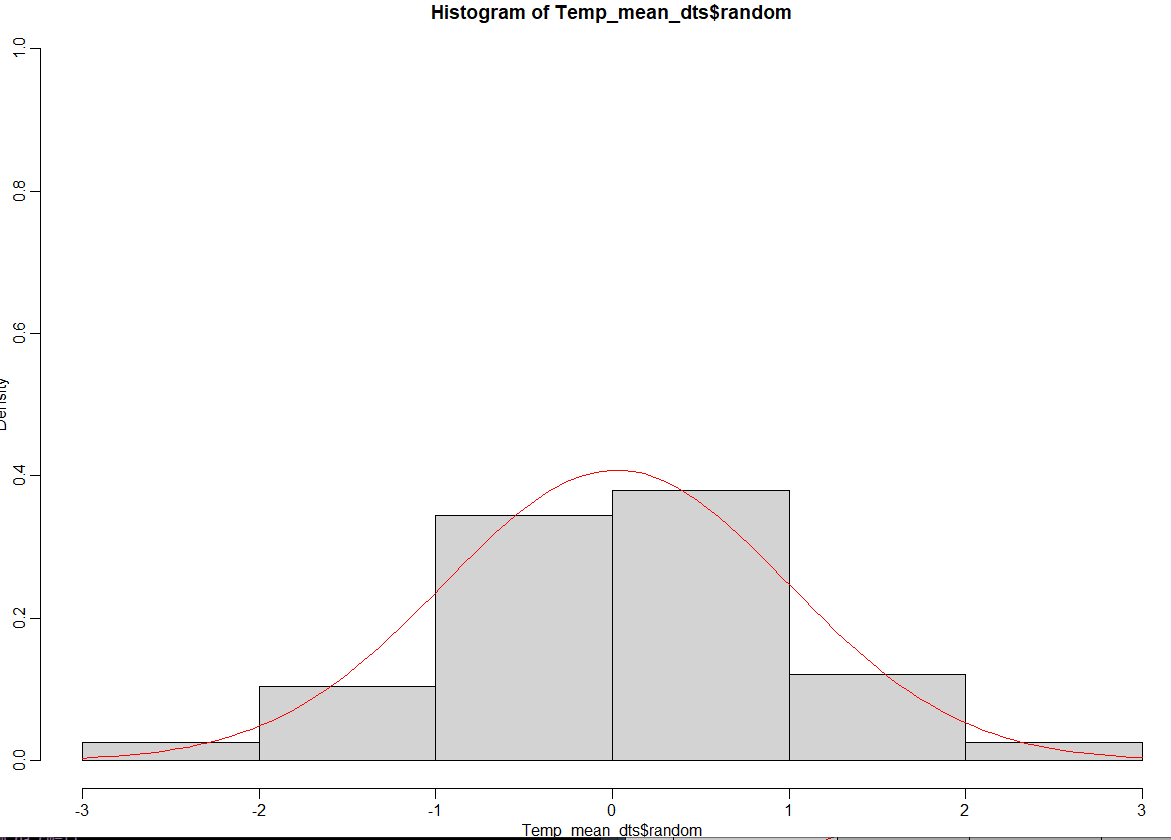
**结果截图：**



**2.2：**使用decompose函数对时间序列进行分解，检查分解后的图像，提取出分解信息中的噪声信息，绘制其直方图及PDF，可以看出噪声满足高斯白噪声的正态分布情况

**结果截图：**

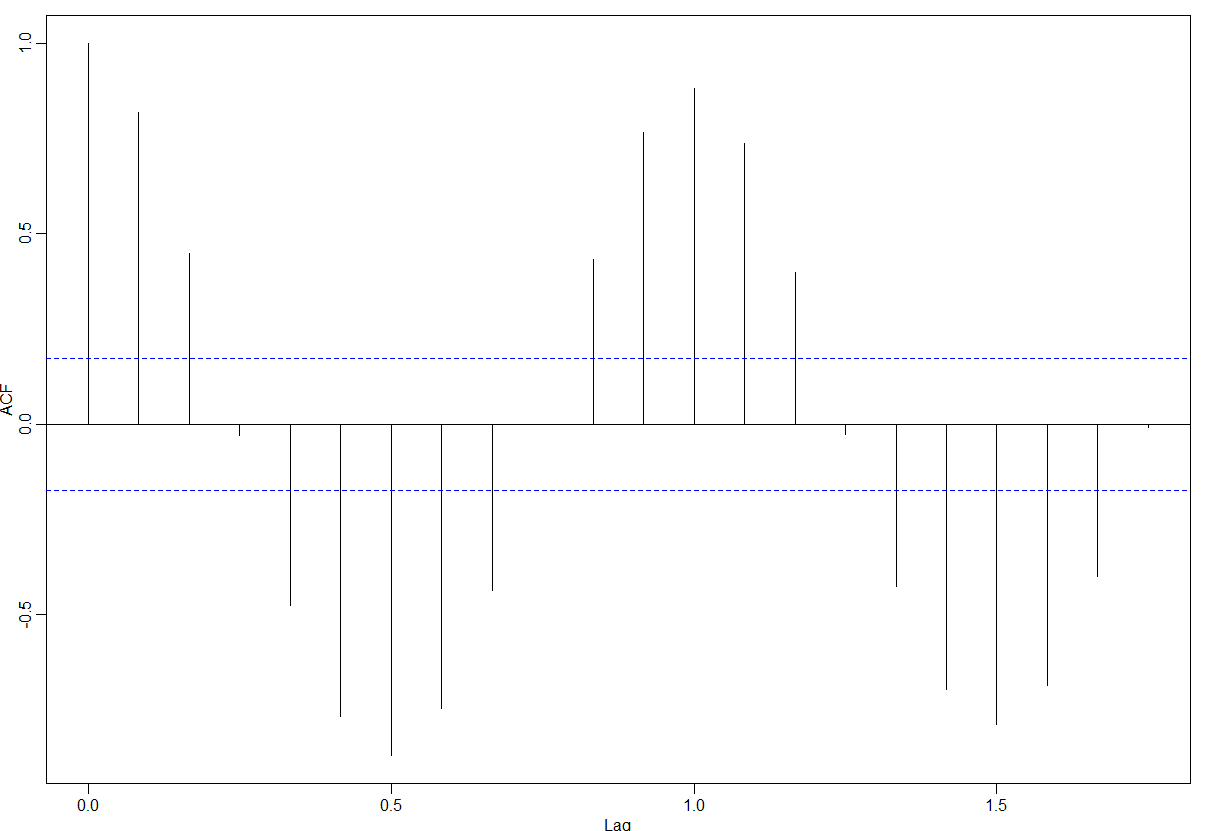


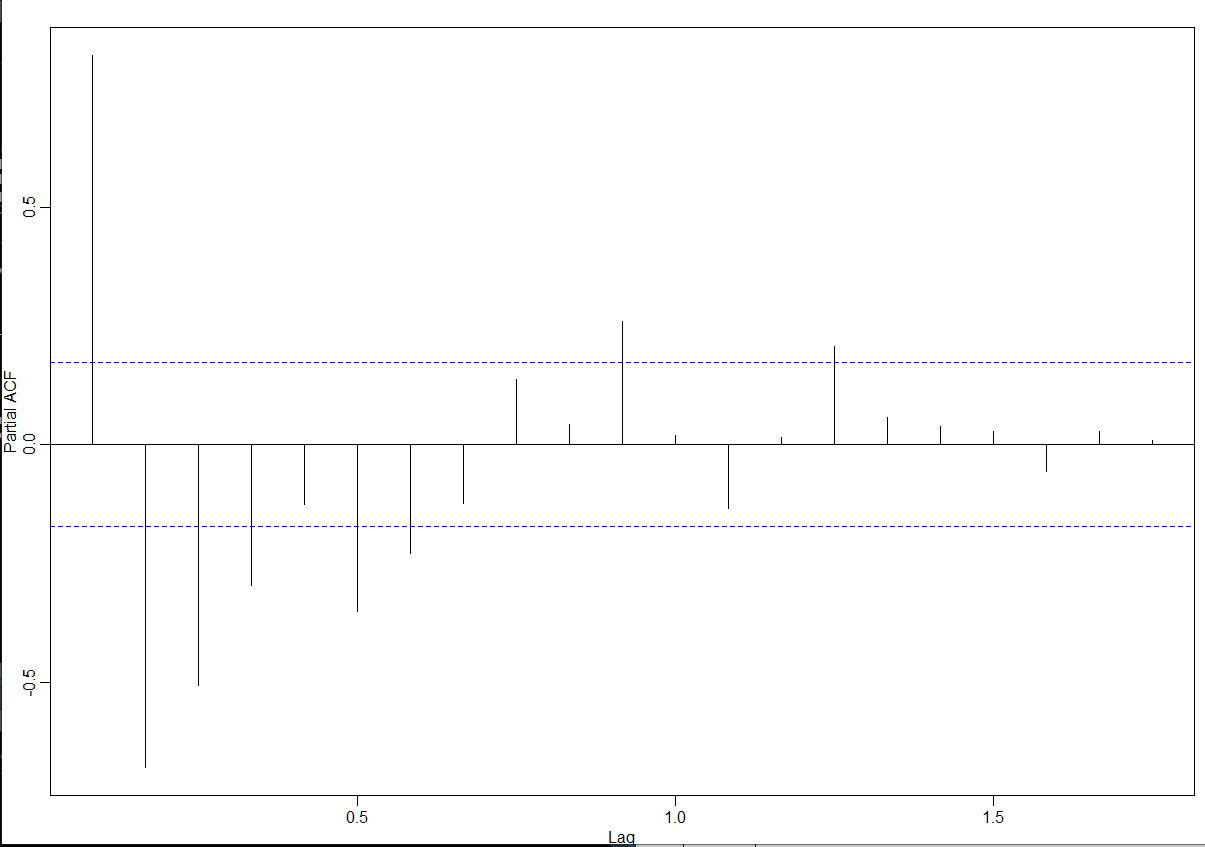


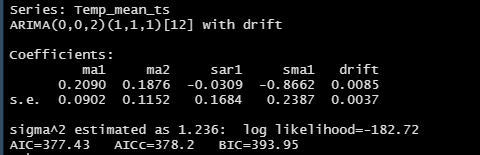
**2.3：拟合过程，**检查月平均温度时间序列的acf和pacf，可以看出acf具有很明显的周期性，结合分解得到的season曲线，可以明显看出受季节影响性很大，常规的arima需要去季节性后才能拟合，由于之后需要带季节预测，这里考虑采用sarima模型（避免要同时预测季节性影响再去叠加arima模型的预测值），代码中采用了两种方式，一种是直接使用auto.arima进行拟合（通过验证发现该函数可以直接求解出sarima模型，即(p,d,q)(P,D,Q)[s]），第二种方式是利用astsa包中的sarima函数进行分析，但具体根据计算情况求解P、D、Q需要结合AIC来优化。

**结果截图：**

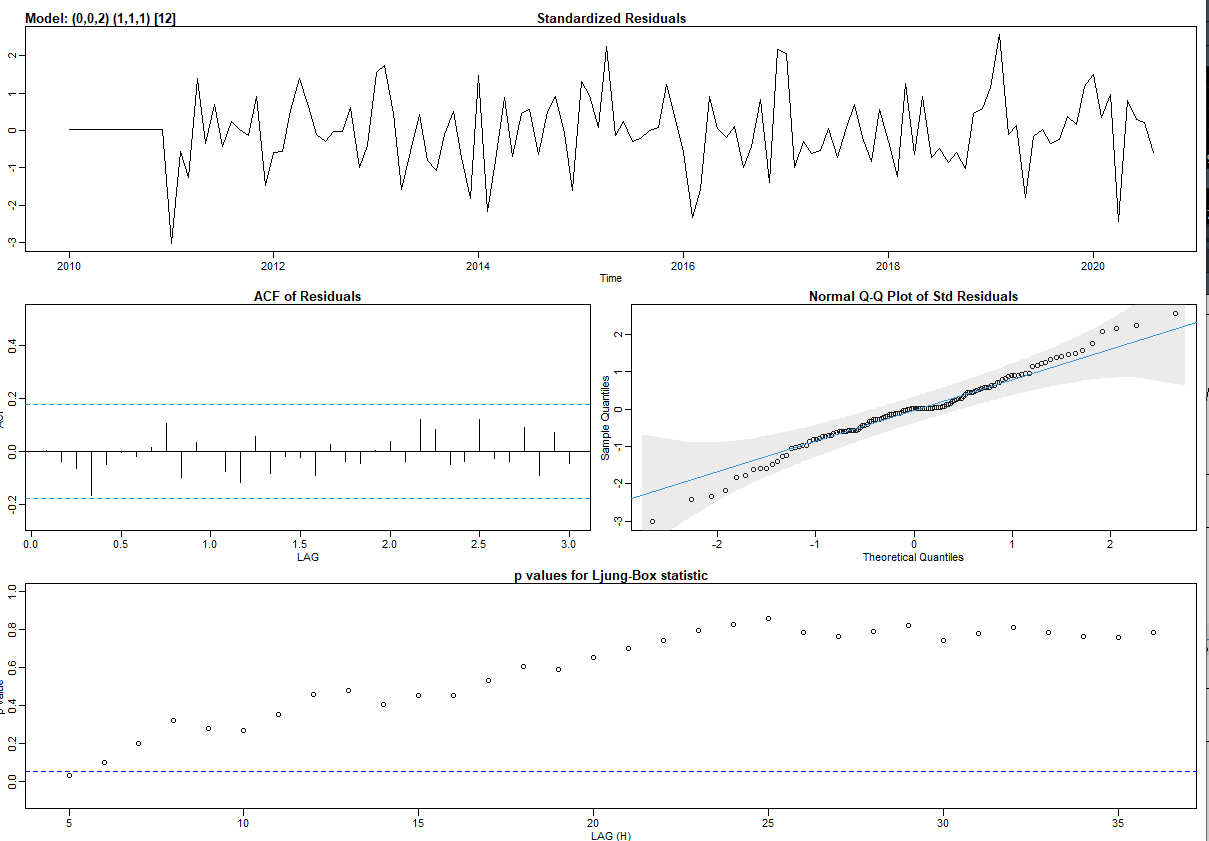
**方法一：**

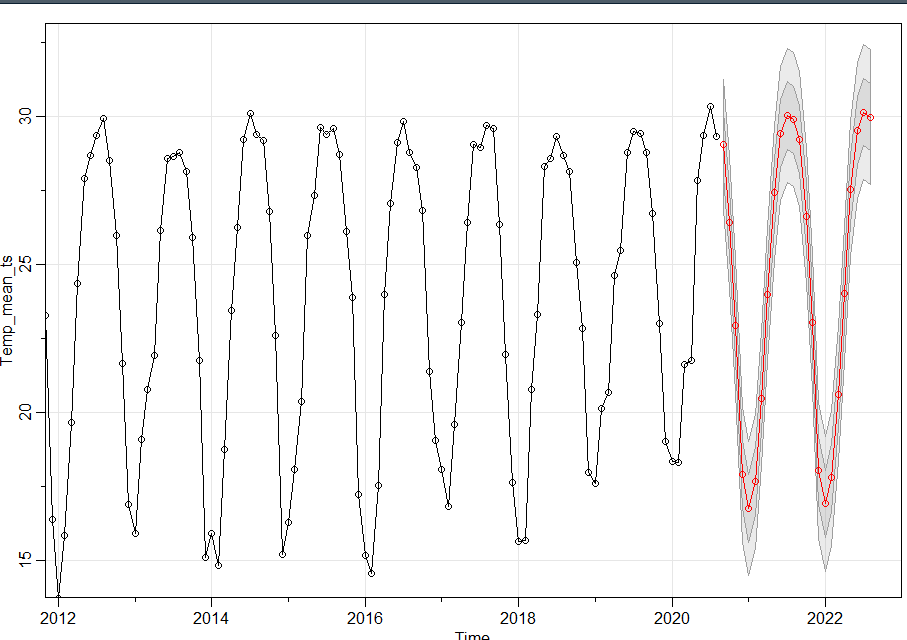






**方法二**





**2.4：**使用建立好的sarima模型对样本进行预测，预测后十个月的平均气温，并求解2020年9月和10月的预测温度及置信区间，分别为9月：29.04255℃，[27.60716，30.47794]；10月：26.40132℃，[24.93491，27.86773]。获取实际9月平均气温29.5摄氏度，十月平均气温从网上获取为23度最低温度，30度最高温度，平均为26.5℃，均位于预测置信区间中，预测良好。

**结果截图：**

