

1.1

使用基于 `ggplot2` 的可视化包 `ggpubr` 的 `ggboxplot` 函数来绘制 `boxplot` 图

首先，加载数据集，输入原始数据

再对如下参数进行调整：

`fill` 填充颜色

`palette` 设置填充的颜色，其中 `npg` 为杂志 `nature` 的配色

`width` 设置箱形图的宽度

`size` 调整轮廓和点的大小

`xlab,ylab` 设置 x 轴、y 轴的标签

`font.x,font.y` 设置 x 轴、y 轴标签的大小

`font.xtickslab,font.ytickslab` 设置 x 轴、y 轴坐标的大小

自定义 x 轴的标签位置

设置主题去掉图例，设置标题的大小、位置等

最后添加统计层，判断这三组样本之间是否显著差异

1.2

对 `economics` 数据集进行时间序列画图

`x=date,y=unemploy`

添加面积图

调整透明度

添加各种标题

换主题

曲线拟合

1.3

正态分布生成 5000 个点

通过分组添加各种各样的颜色

去掉图例

调整色调，色彩饱和度和亮度

绘制概率密度函数（`rstudio` 中不知道为啥没有出现）

调整主题

1.4

以 `iris` 数据集为例，绘制花萼的长度和宽度的关系

`color = Species`

`size = 2.0`

`shape = 16`

增加了 `labs`

还有 `facet_wrap(~ Species)` 分成几个图

再对花萼长度与宽度进行拟合

用 `ggthemes` 这个包修改 `theme`

1.5

课后习题第二题

按要求画出气温异常的地方

用了超多的参数以及变量

可以从图中看出东经 100 度赤道地区气温存在异常

2.1

首先输入数据

转 tibble

用 mutate 和 ifelse 删选初始数据得到气温

再用切片函数和粘贴函数得到月份

再筛选 201001-202008

再 groupby 和 summarise

得到每月平均气温

用 ts 函数转时间序列并画图

2.2

用 components 分解, 提取 random

再用 Box-Ljung test 检验

得到 p 值为 $1.175e-05$

p 值小于 0.05, 说明接受为白噪声

2.3

先做差分检查

再用 acf 和 pacf 进行检验

最后用 auto.arima 自动建模

ARIMA(0,0,2)(1,1,1)[12] with drift

2.4

用 forecast 预测未来 2 个月的天气

再取 9 月和 10 月的平均值

最后提取出 9 月真值进行比较, 数据没有 10 月的数据。

相对偏差不到 1.4%

预测还是挺成功的