

#### 4.1.1

使用基于 ggplot2 的可视化的 boxplot 函数来绘制 boxplot 图

首先加载数据集 midwest,

再对生成图像参数进行调整：

varwidth=T

fill 填充黄色

标题为"Box plot",

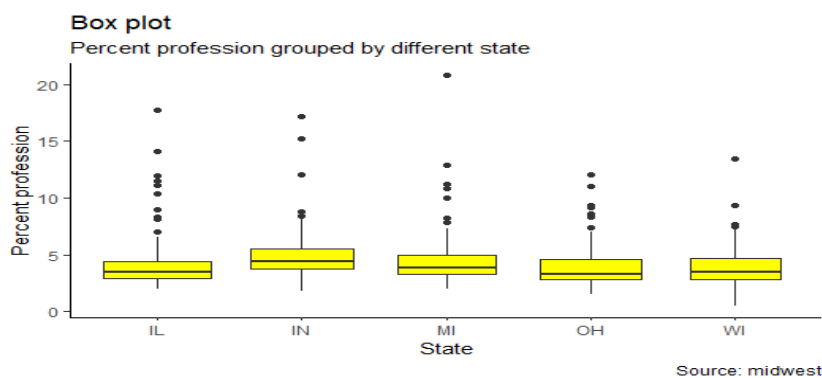
次标题设置为"Percent profession grouped by different state",

caption="Source: midwest",

设置 x 轴、y 轴的标签

x="State",

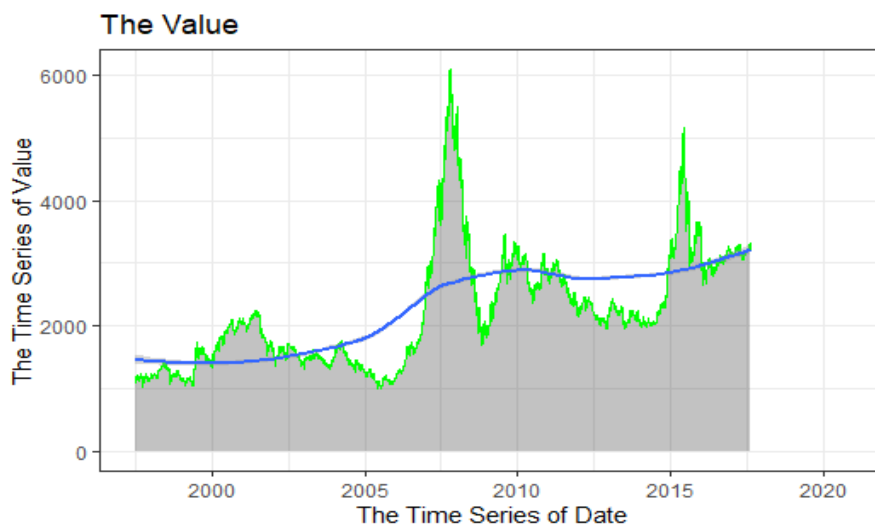
y="Percent profession"



#### 4.1.2

使用基于 ggplot2 的可视化函数来绘图，数据使用定量金融模型框架 quantmod。

首先利用 quantmod 包从 yahoo 数据源获取从 1997 年以来的数据，存于变量 SSEC 中，抽取收盘数字，然后分别提取时间数据和指数数值，绘图结果如下图



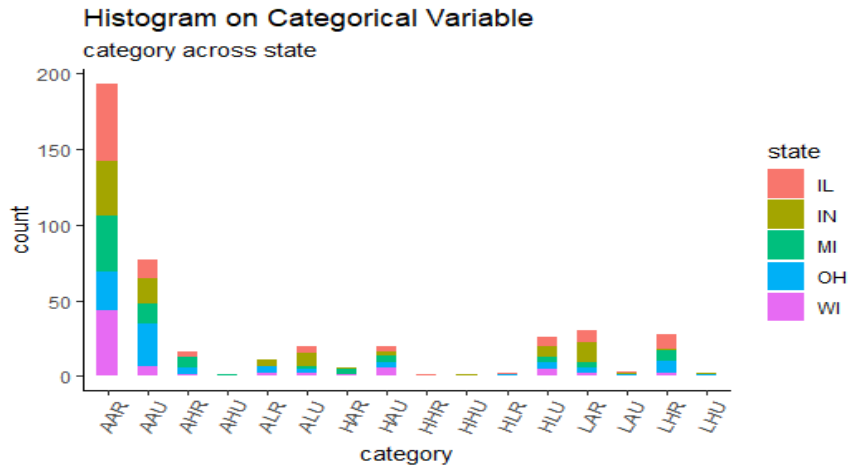
#### 4.1.3

频率直方图显示每个类别的条形图。

通过调整 width，您可以调整条形方块的宽度。

数据采自美国中西部县人口资料数据包 midwest。

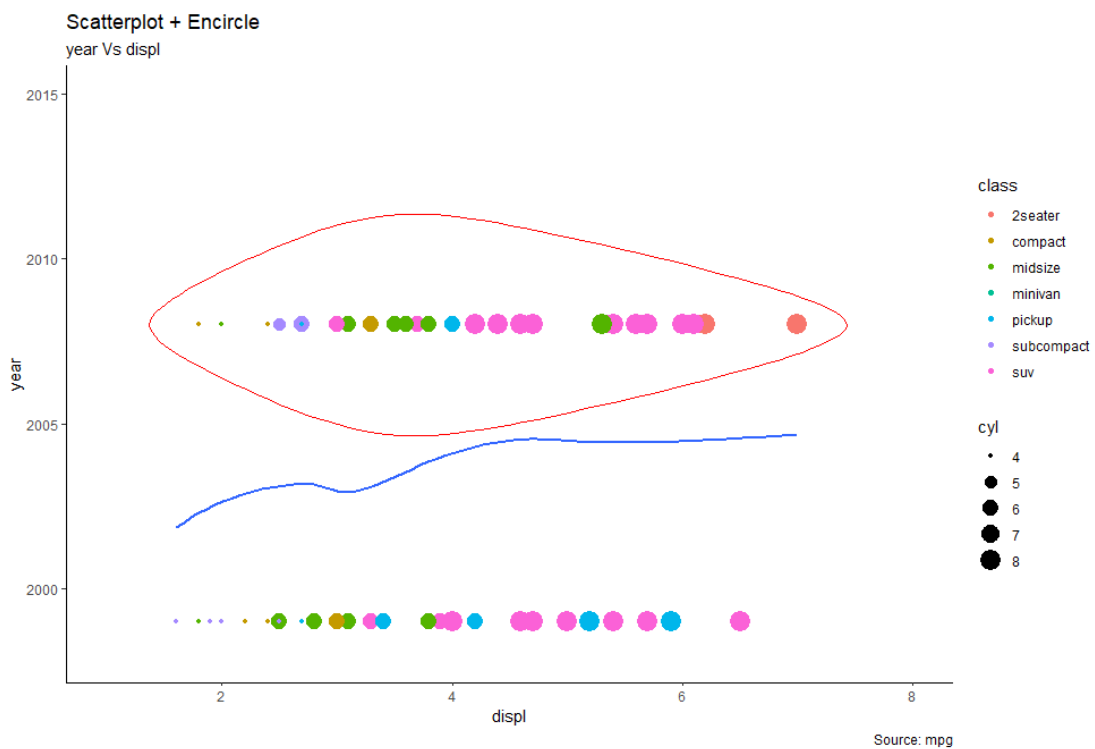
使用基于 ggplot2 的可视化函数来绘制分类变量的直方图。



#### 4.1.4

在介绍结果时, 在图表中添加某些特殊的点或区域组, 以便引起人们对这些特殊情况的注意。使用 `ggalt` 包中的 `geom_encircle()` 可以方便地完成此操作。

在中 `geom_encircle()`, 将 `data` 设置为仅包含点(行)或兴趣点的新数据框。此外, 使用 `expand` 使曲线通过点外。曲线的 `color` 和 `size` (厚度) 也可以修改。



上图的绘制采用的数据集为 38 款流行车型 1999 年至 2008 年的燃油经济性数据

上图主要是为了说明不同品牌的汽车在不同年代生产的不同排量的汽车。

#### 4.1.5

使用基于 `ggplot2` 的可视化函数来绘图, 数据使用价格超过 5 万圆切割钻石数据包 `diamonds`。

通过调整不同的 `position` 绘制不一样的直方图。

使用

`bar + coord_flip()`      `bar + coord_polar()` 画出两种不同的占比图。

#### 4.2.1

读取数据，转换 tibble 数据类型。用 mutate 和 ifelse 清洗初始数据。

再用 substr 函数和 paste 函数重新制作一个日期格式。

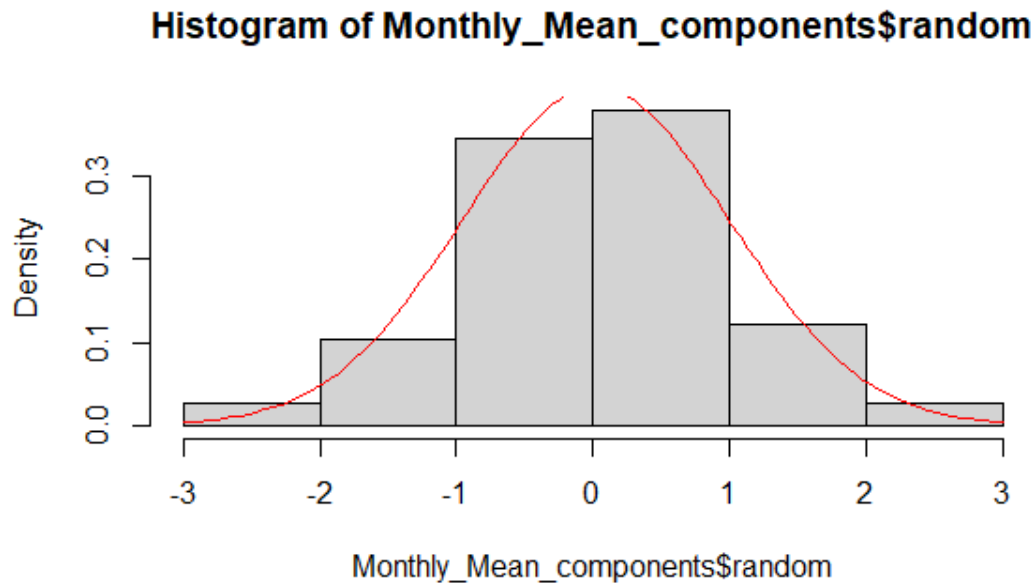
清洗后的温度数据再筛选 得到 201001-202008 的数据。

使用 groupby 和 summarise 得到每月平均气温。使用 ts 函数转时间序列并做图。

#### 4.2.2

用 components 分解时间序列得到 random，检查错误部分是否遵循白噪声分布

plot hist & add pdf,观察绘制的曲线，符合正态分布，说明接受为白噪声。



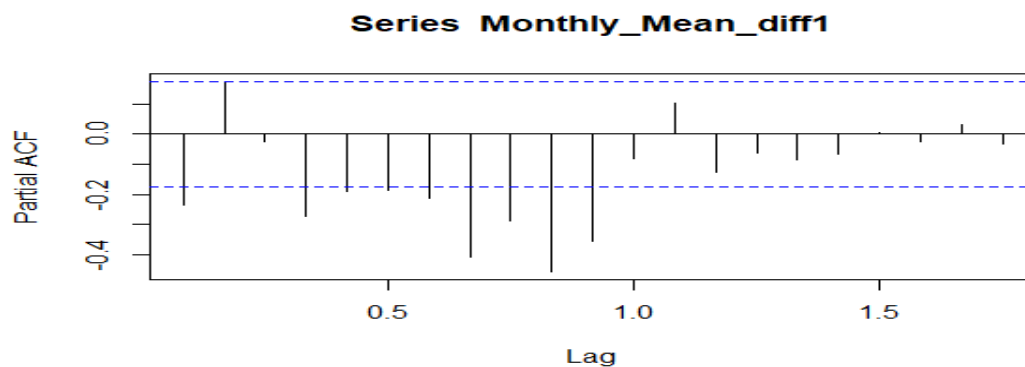
#### 4.2.3

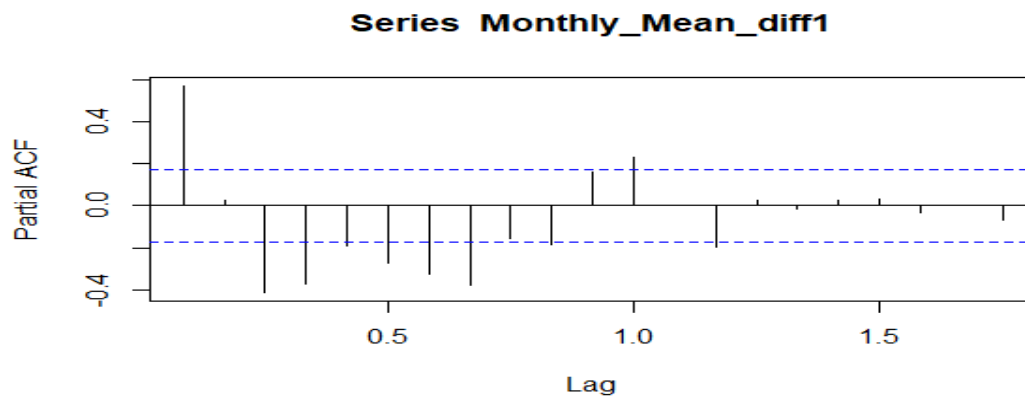
拟合一个 arima 模型，先做差分检查。

画出相关图，再用 acf 和 pacf 进行检验，绘制部分相关图。

最后用 auto.arima 自动建模

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





#### 4.2.5

用 ARIMA 模型预测每月平均气温，拟合一个  $(3,1,1)$  的模型。

用 `acf` 和 `box.test` 进行检验。

绘制预测误差的时间图。

根据拟合结果图和实际数据对比可看出预测成功。

