《气象台数据的爬取与分析》报告书

姓名：董欣然

学号：1900013018

**一.概述**

为了直观地了解中国近年的气候变化，我爬取了[气象网站](http://www.tianqihoubao.com/)2011年至2020年十年来，中国主要城市(省会城市)的气象数据，总计443,520条。

我使用数据库sqlite存储数据，使用pandas进行数据的统计分析，使用pycharts等工具将数据可视化。通过充分地挖掘和利用数据，我们可以轻松直观地了解到国内气候的变化，了解气候与省份、年份、月份之间的内在联系。

文档不能完全展示我的工作成果，请浏览**”报告书网页版.html”**和”**报告书视频版.mp4**”（在web文件夹中）了解详情。

**二.成果展示**

1. **数据爬取**

网址： <http://www.tianqihoubao.com/>

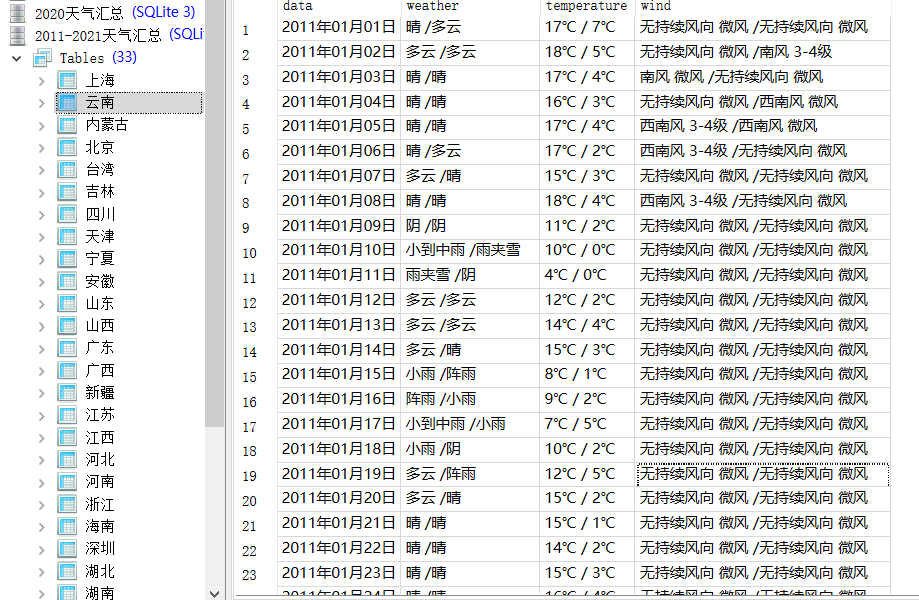
成果：爬取了从2011年至2021年中国省会城市每日的天气信息。信息包括每日的气象、风向、风强、最高温、最低温等。数据信息总计443,520条，已存储入数据库中。

1. **数据库存储**

每个省份单独建立table，将该省份的气象信息储存在其中。

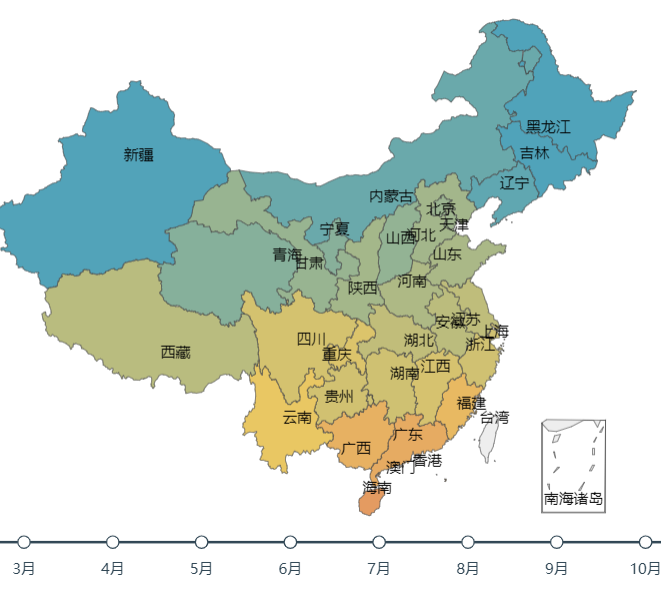
每个table有4个字段，分别对应着日期、气象、最高温以及最低温、风向以及风强。

数据库缩略图如下：



1. **中国气温地图**

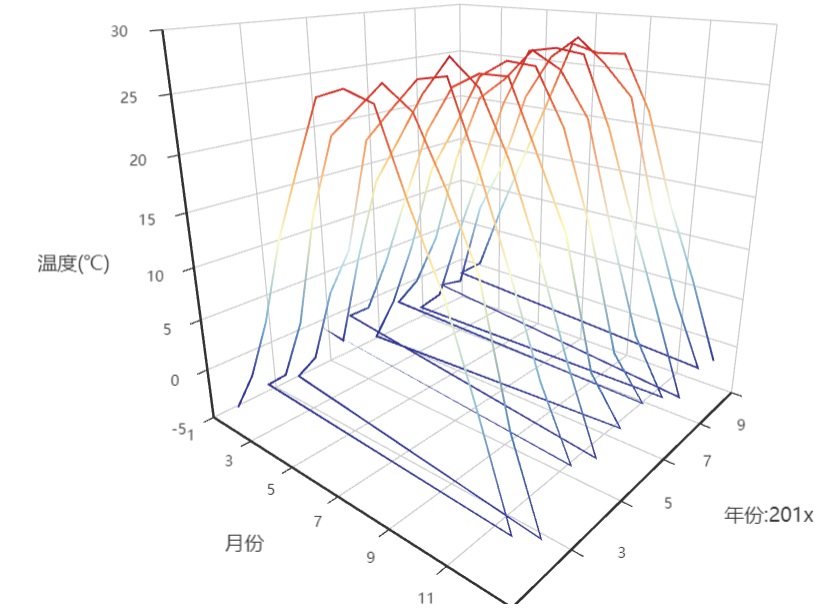
我对近十年来各大城市总计221,760条气温数据进行了统计，整合出从2011年至2021年共10张气温地图，每张地图又由从1月至12月共12张气温地图按时间轴组合在一起。我采用色彩的变化表示温度的变化，从而可以直观地看出各地区的气温差异、各月份的气温变化。缩略图如下



1. **年、月份、温度三维立体图**

我对各地区的气候变化展开进一步分析，绘制了**年、月份、温度三维立体图。**3D折线图以年份为x轴，月份为y轴，气温为z轴，既展现了一年中气温的变化，又展现了不同年份之间的气温差异。

缩略图如下：

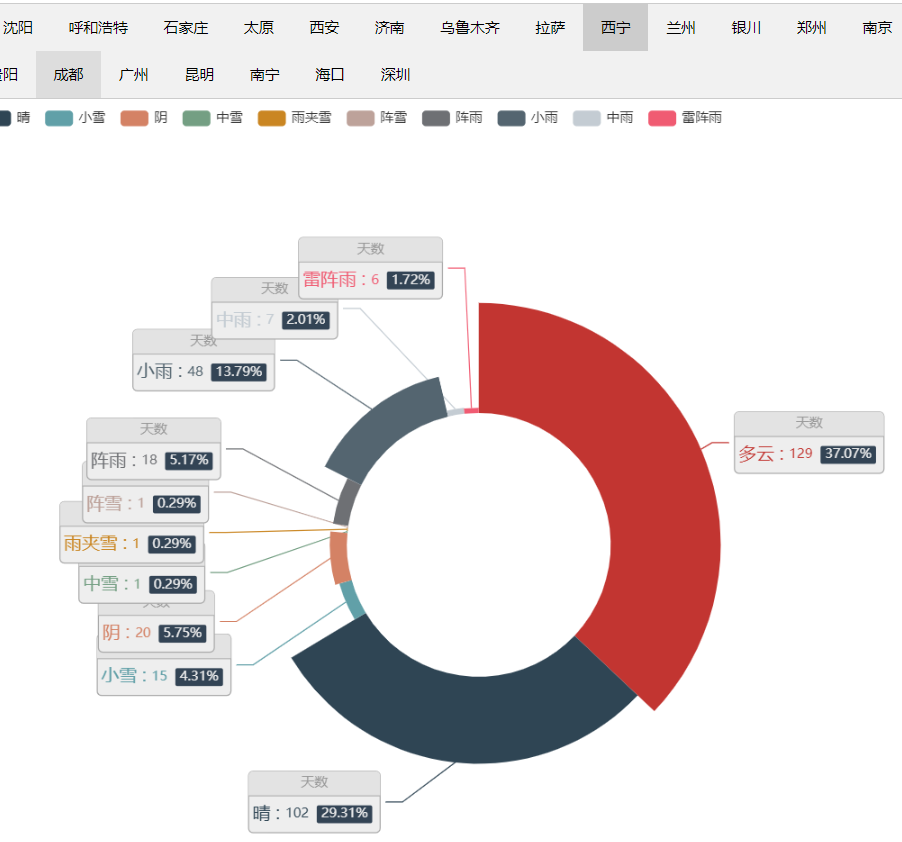


1. **晴雨表**

气温不能完全刻画气候特征，我们从气象的角度进一步分析各地区间的气候特点。

通过统计计算，我绘制了33张晴雨表。晴雨表展示了中国主要城市全年各种气象出现的频次以及在全年中的占比。

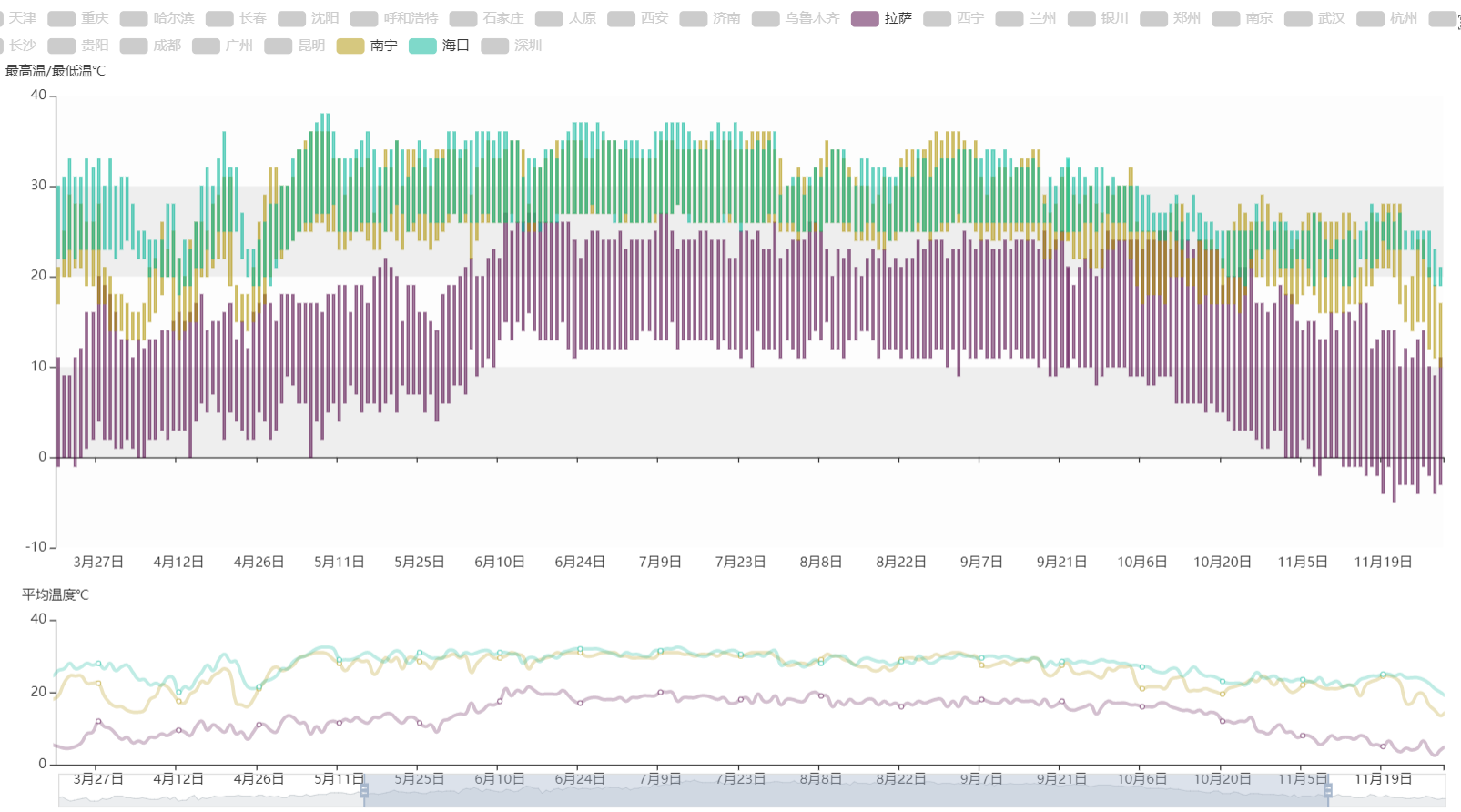
缩略图如下：



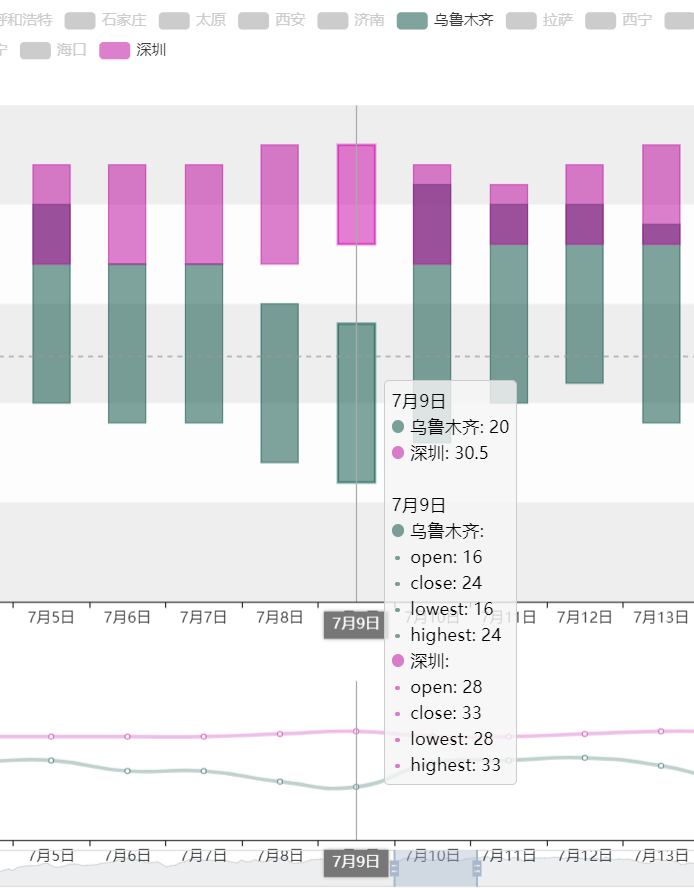
1. **全年最高(低)温度柱状图、平均温度折线图**

一天内的温度有高低变化，不同地域的昼夜温差也不尽相同。平均温度不能完全刻画中国气候特征，故我绘制了**全年最高(低)温度柱状图、平均温度折线图**。缩略图如下。

* 图中的柱状部分表示一天内的温度变化区间。
* 折线图中的顶点表示当天的平均气温。
* 图中顶部是选项卡，可以查看任意几个城市的气温变化。
* 图中底部是可拖拽的时间轴。可以详细查看某一时间段的气温详情（时间的最小精度是一天）。



上图选择的城市是拉萨、南宁、海口，时间区间是2020年5月14日至2020年11月8日。可以看出南宁、海口的气候较为相近，拉萨气温低、昼夜温差大。



上图选择的城市是乌鲁木齐、深圳，时间区间是2020年7月5日至2020年7月13日。用鼠标轻触时间轴（如图显示的是7月9日）会浮现标签页，标签页上标明了城市当天的气温详情。

**三.总结**

我使用了一系列python工具，对气象台数据进行了爬取、存储、清理、计算、可视化。最终绘制了五类图表：

* 中国气象地图
* 年、月份、温度三维立体图
* 晴雨表
* 全年最高(低)温度柱状图
* 全年平均温度折线图

为了充分地展现数据，提高用户使用的体验，每类图表又包含选项卡、功能卡、多张子图。 文档不能完全展示我的工作成果，请查看**”报告书网页版.html”**和”**报告书视频版.mp4**”（在web文件夹中）进一步了解。