

《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](#) 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。

2. 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。

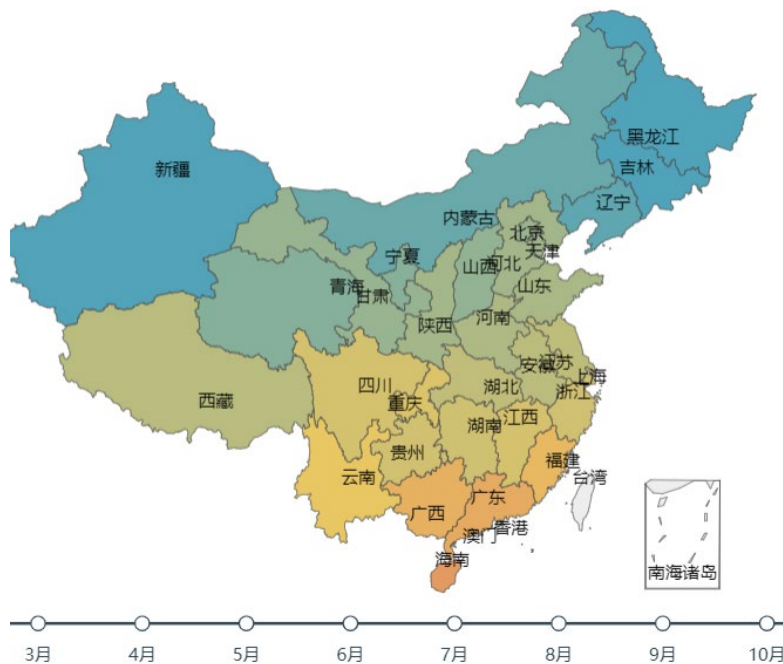
每个 table 有 4 个字段，分别对应着日期、气象、最高温以及最低温、风向以及风强。

数据库缩略图如下：

	data	weather	temperature	wind
1	2011年01月01日	晴/多云	17°C / 7°C	无持续风向 微风 / 无持续风向 微风
2	2011年01月02日	多云/多云	18°C / 5°C	无持续风向 微风 / 南风 3-4级
3	2011年01月03日	晴/晴	17°C / 4°C	南风 微风 / 无持续风向 微风
4	2011年01月04日	晴/晴	16°C / 3°C	无持续风向 微风 / 西南风 微风
5	2011年01月05日	晴/晴	17°C / 4°C	西南风 3-4级 / 西南风 微风
6	2011年01月06日	晴/多云	17°C / 2°C	西南风 3-4级 / 无持续风向 微风
7	2011年01月07日	多云/晴	15°C / 3°C	无持续风向 微风 / 无持续风向 微风
8	2011年01月08日	晴/晴	18°C / 4°C	西南风 3-4级 / 无持续风向 微风
9	2011年01月09日	阴/阴	11°C / 2°C	无持续风向 微风 / 无持续风向 微风
10	2011年01月10日	小到中雨/雨夹雪	10°C / 0°C	无持续风向 微风 / 无持续风向 微风
11	2011年01月11日	雨夹雪/阴	4°C / 0°C	无持续风向 微风 / 无持续风向 微风
12	2011年01月12日	多云/多云	12°C / 2°C	无持续风向 微风 / 无持续风向 微风
13	2011年01月13日	多云/多云	14°C / 4°C	无持续风向 微风 / 无持续风向 微风
14	2011年01月14日	多云/晴	15°C / 3°C	无持续风向 微风 / 无持续风向 微风
15	2011年01月15日	小雨/阵雨	8°C / 1°C	无持续风向 微风 / 无持续风向 微风
16	2011年01月16日	阵雨/小雨	9°C / 2°C	无持续风向 微风 / 无持续风向 微风
17	2011年01月17日	小到中雨/小雨	7°C / 5°C	无持续风向 微风 / 无持续风向 微风
18	2011年01月18日	小雨/阴	10°C / 2°C	无持续风向 微风 / 无持续风向 微风
19	2011年01月19日	多云/阵雨	12°C / 5°C	无持续风向 微风 / 无持续风向 微风
20	2011年01月20日	多云/晴	15°C / 2°C	无持续风向 微风 / 无持续风向 微风
21	2011年01月21日	晴/晴	15°C / 1°C	无持续风向 微风 / 无持续风向 微风
22	2011年01月22日	晴/晴	14°C / 2°C	无持续风向 微风 / 无持续风向 微风
23	2011年01月23日	晴/晴	15°C / 3°C	无持续风向 微风 / 无持续风向 微风

3. 中国气温地图

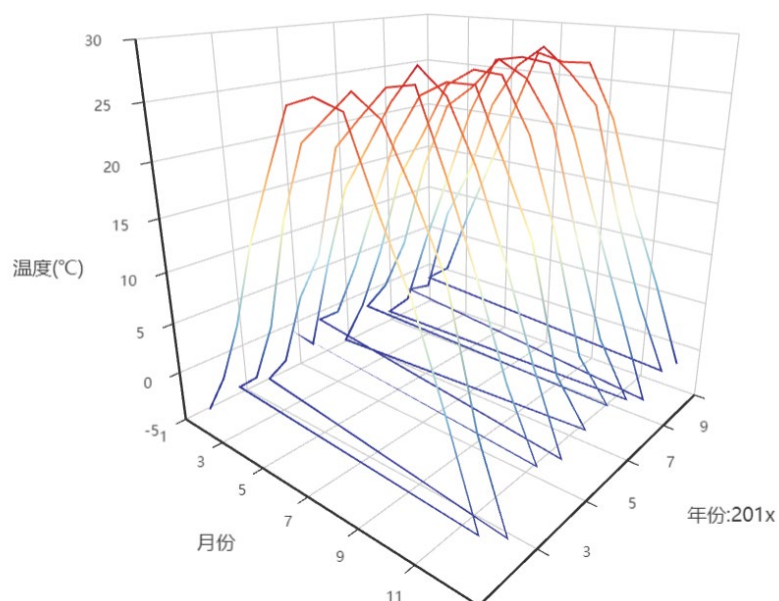
我对近十年来各大城市总计 221,760 条气温数据进行了统计，整合出从 2011 年至 2021 年共 10 张气温地图，每张地图又由从 1 月至 12 月共 12 张气温地图按时间轴组合在一起。我采用色彩的变化表示温度的变化，从而可以直观地看出各地区的气温差异、各月份的气温变化。缩略图如下



4. 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：

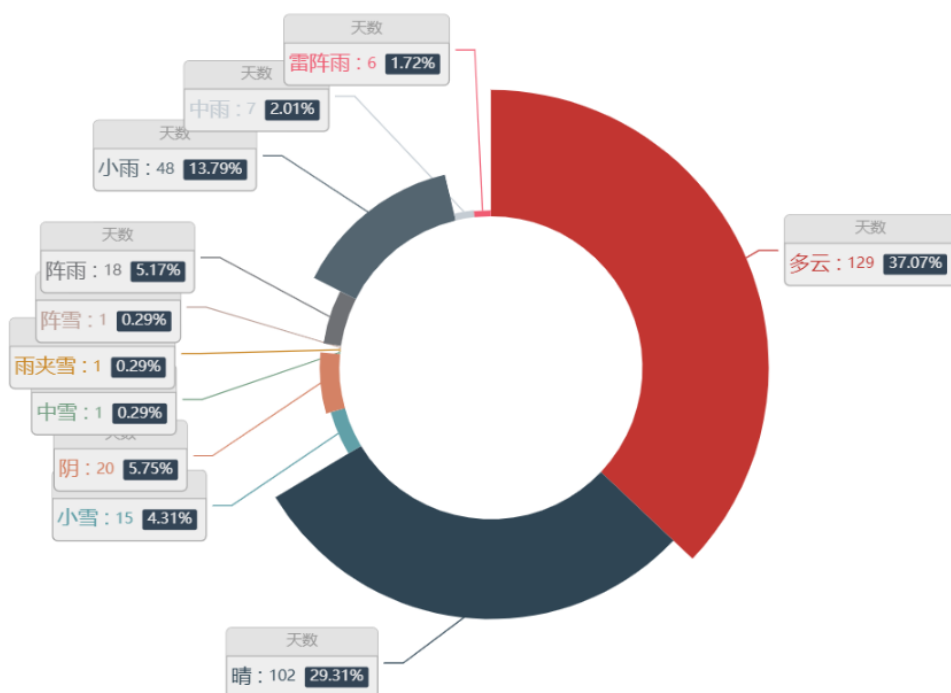
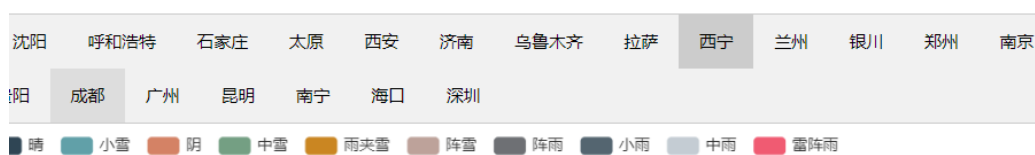


5. 晴雨表

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。

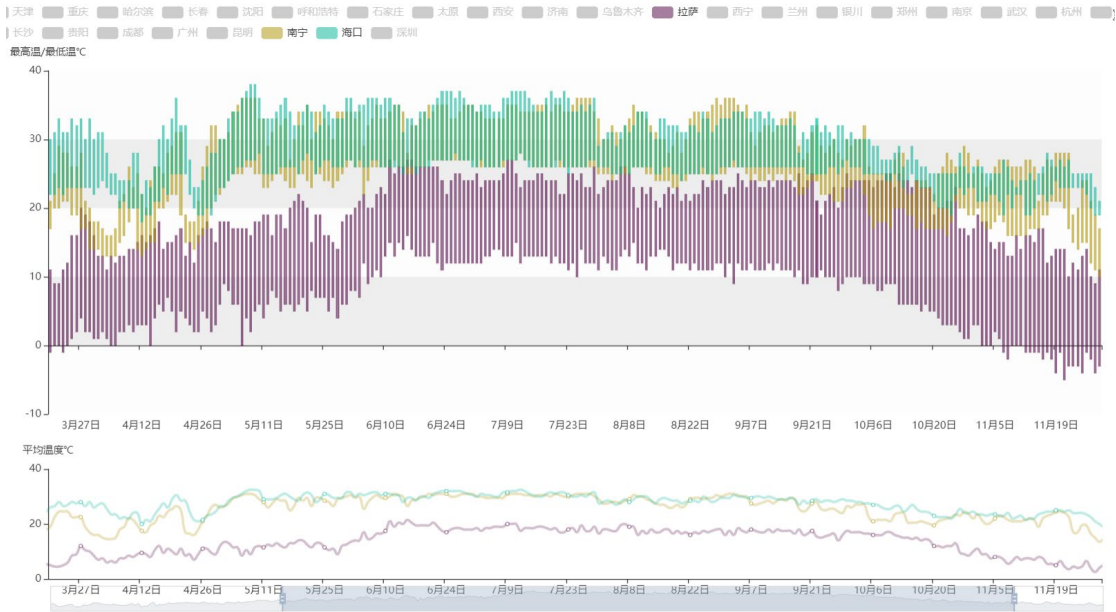
缩略图如下：



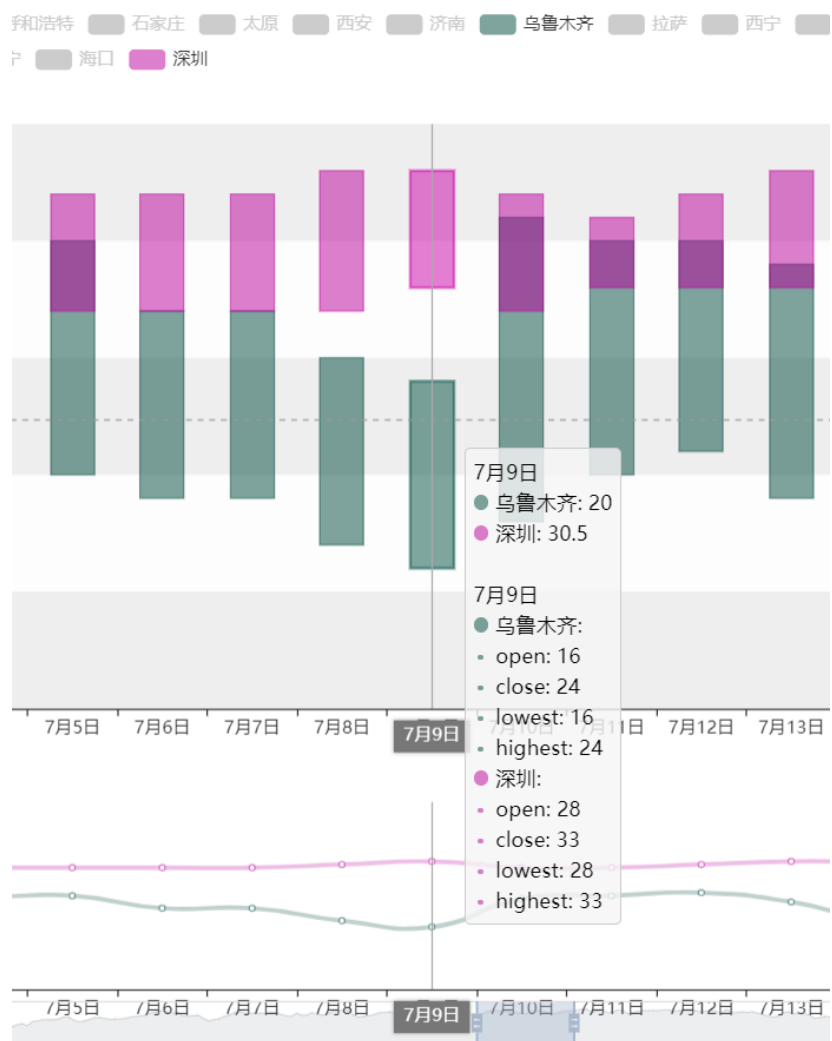
6. 全年最高(低)温度柱状图、平均温度折线图

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。