复习题

1. **数据采集，数据存储，数据管理，数据计算，数据分析，数据展示**
2. **数据采集是指从真实世界对象中获得原始数据的过程，不准确的数据采集将影响后续的数据处理并最终得到无效的结果。**目前，最常用的三种数据采集方法是，传感器，日志文件和爬虫
3. **数据管理是利用计算机硬件和软件技术对数据进行有效的收集，存储，处理和应用的过程。大数据管理的本质是在大数据基础设施层之上向用户提供更加方便，高效，友好的人数交互界面，随着大数据处理技术的发展，大数据的终端用户将越来越不用关心底层的基础设施，计算，存储，网络，细节，只用关心数据本身，以及如何来开展大数据应用的创新与服务**
4. 批量计算，流式计算，交互式计算和图计算模式
5. **数据分析处理来自对某一兴趣现象的观察，测量或者实验的信息。数据分析目的是从和主题相关的数据中提取尽可能多的信息。方法有统计分析，数据挖掘，机器学习和可视化分析等**
6. **数据可视化的应用价值 ,其多样性和表现力吸引了许多使用者 ,无论是动态还是静态的可视化图形 ,都为我们搭建了新的桥梁 ,让我们能洞察世界的究竟 ,发现形形色色的关系 ,感受每时每刻围绕在我们身边的信息变化 ,还能让我们理解其他形式下不易发掘的事物 。**

**践习题**

**7，8代码和图片见Sept.30文件夹**

**7的数据集来源于大学生统计竞赛的烂尾项目(x)**

**8是官方的数据集tips**