14 数据ETL作业

- 提供的资料：
  - 参考2个国家标准《GB 11643-1999 公民身份证号码》、《GB/T 2260-2007 中华人民共和国行政区划代码 》
  - data文件： test_id_card_no.csv
- 1、分别查询：
  - 吉林省的身份证号码（参考GB/T2260-2007标准）
  - 校验码不正确的无效身份证号码
  - 50岁以上人的身份证号码
- 2、分别排序：
  - 按年龄降序排列
  - 先按性别排序（女士在前），若性别相同则按年龄升序排列
- 3、生成新列：
  - 性别（列名sex，类型为factor，值为Female或Male）、生日（列名Birthday，类型为date）、有效否（列名Valid，类型Logic）

14 数据ETL作业

- 4、统计：
  - 所有人的平均年龄（允许小数）、年龄的中位数
  - 总人数
  - 是否有30岁以内的人
- 5、分组汇总：
  - 男女人数、男女平均年龄
  - 男女年龄超过50岁的人数和比例
- 6、综合：
  - （1）将同时满足下面几个条件：校验码有效、65岁以内（含65）、非长春市人、按年龄降序排列的Id_Card_No、sex、Birthday、Valid这4个列的数据保存为Out_Id_Card_Data.csv文件；
  - （2）同时按省级行政区划和性别分组，然后对每组统计老人和非老人的人数和比例（>=60为老人）