b

学号 ： 20194169 姓名： 麻海志 等级：

设计要求：请应用Tableau Desktop软件分析“120年奥运史情况”，要求其中至少包含5种可视化分析图形和一组仪表板（含故事），并予以发布（打印）。

数据来源：Kaggle"120 years of Olympic history: athletes and results"数据集。

样本数据：包括athlete\_events.csv和noc\_regions.csv两个文件，有从1896年雅典

奥运会到2016年里约奥运会的每一届奥运会的数据。

栏目说明：athlete\_events.csv文件包含15个字段、271116条记录。每一条记录对应一

位在一项奥运会项目中参赛的运动员。具体字段及对应含义如下：

ID - 运动员ID编号

Name - 运动员姓名

Sex - 性别

Age - 年龄

Height - 身高(cm)

Weight - 体重(kg)

Team - 队伍名称

NOC - 国家奥委会编码

Games - 奥运会年份和季节

Year - 年份

Season - 季节

City - 主办城市

Sport - 体育运动

Event - 比赛项目

Medal - 获奖情况(金、银、铜、未获奖)

noc\_regions.csv描述的是国家奥委会编码与具体的国家(地区)名称的对应关系。包含3

个字段，分别是NOC国家奥委会编码、regions国家(/地区)具体名称及notes备注。

我的记录：

**1.我创建的可视化图标**

(1).

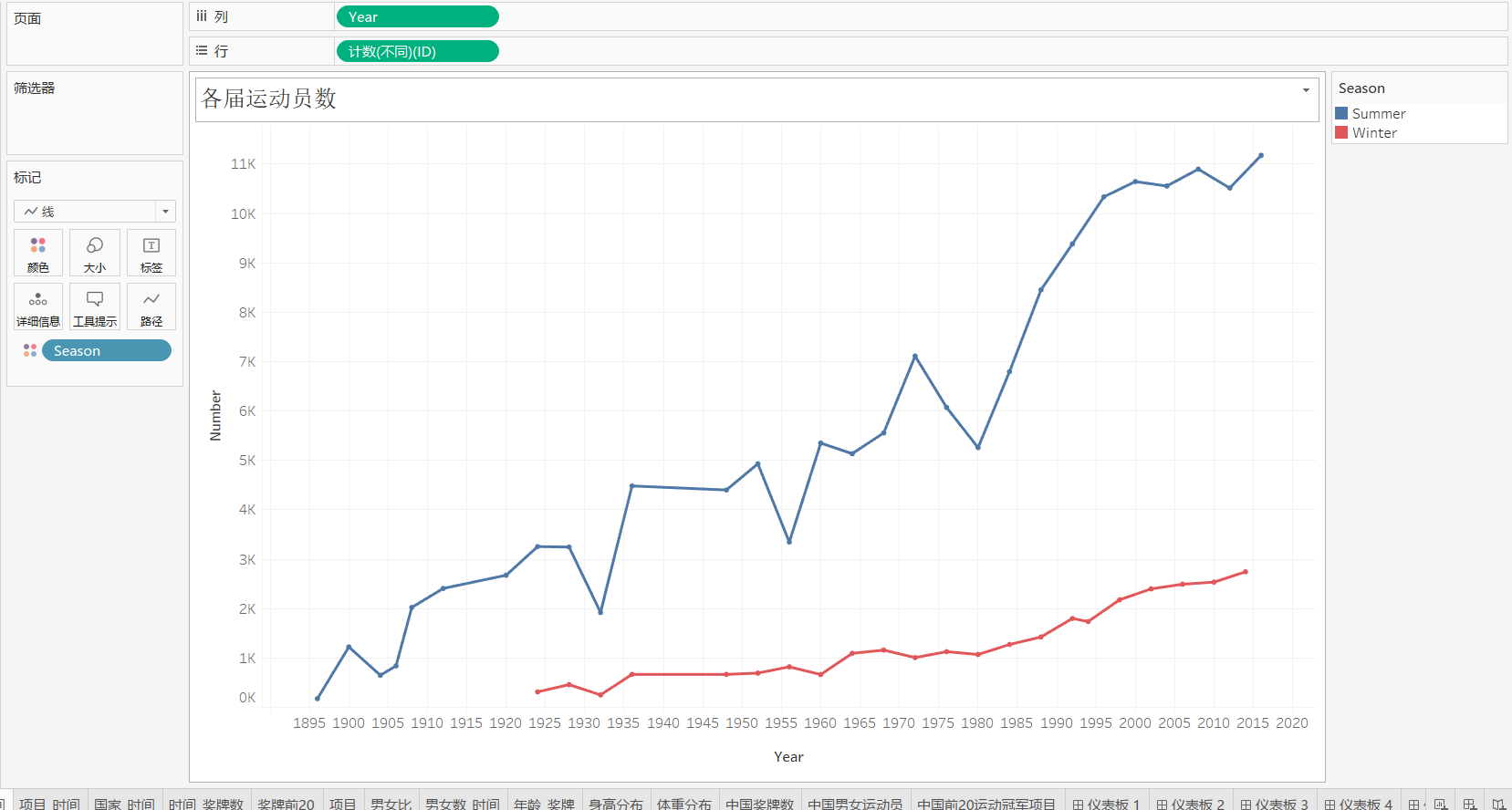
基于运动员的ID和年份字段，来绘制一个各届参赛运动员数变化的折线图。下面将ID的类型改为度量→计数(不同)，Y轴即为每届奥运会运动员的总人数。但由于原文件夏季奥运会和冬季奥运会数据是混合的，故折线图形状明显异常。将Season拖至标记卡中的【颜色】，就能看到夏季和冬季的数据分成了两条折线，每个类型对应一种颜色，并自动生成图例，完成了一幅基本的折线图。对图表标题、颜色、坐标轴标签、坐标轴范围等进行设置，只需点选预览即可。最终效果如下图1.1所示：

图1.1-1

(2).

同理，将Year和Event分别拖到列和行中，Event设置为不同计数，然后再把season拉到筛选器中，绘制各届赛事项目数、各界参与国家数折线图。效果如下图2.1-1所示。接着再创建一个工作表将Year和Region分别拖到列和行中，绘制出各界参与奥运会的国家数量，效果如图2.1-2所示

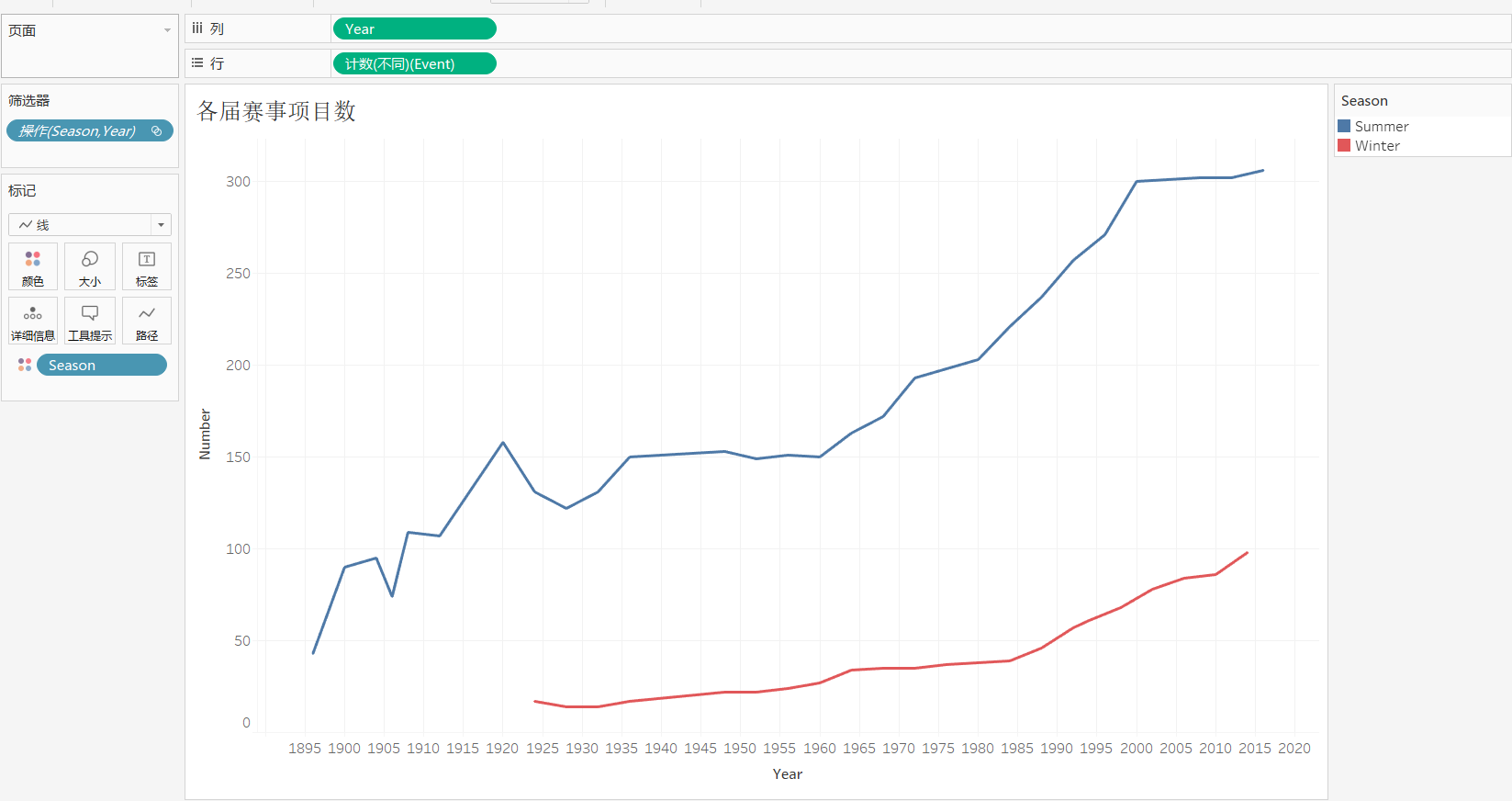
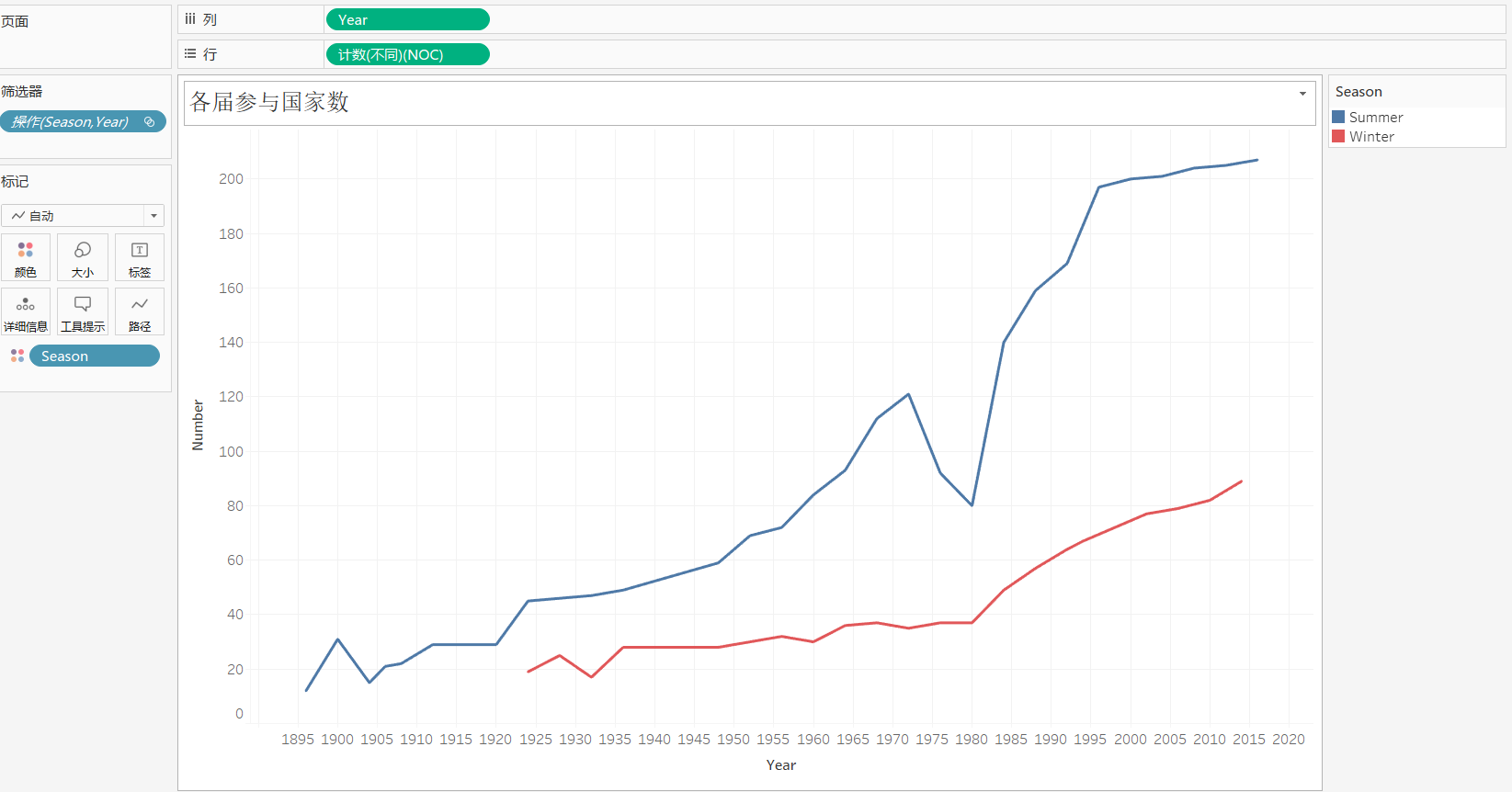


图2.1-1

图2.1-2

(3).

绘制奖牌与年龄的之间的关系图，生成过程与折线图基本相同，只需修改图表类型。在Tableau界面右端有一个【智能推荐】按钮，可快速创建基本图形，光标移动到图形上就会显示字段要求，效果如下图所示3.1-1。

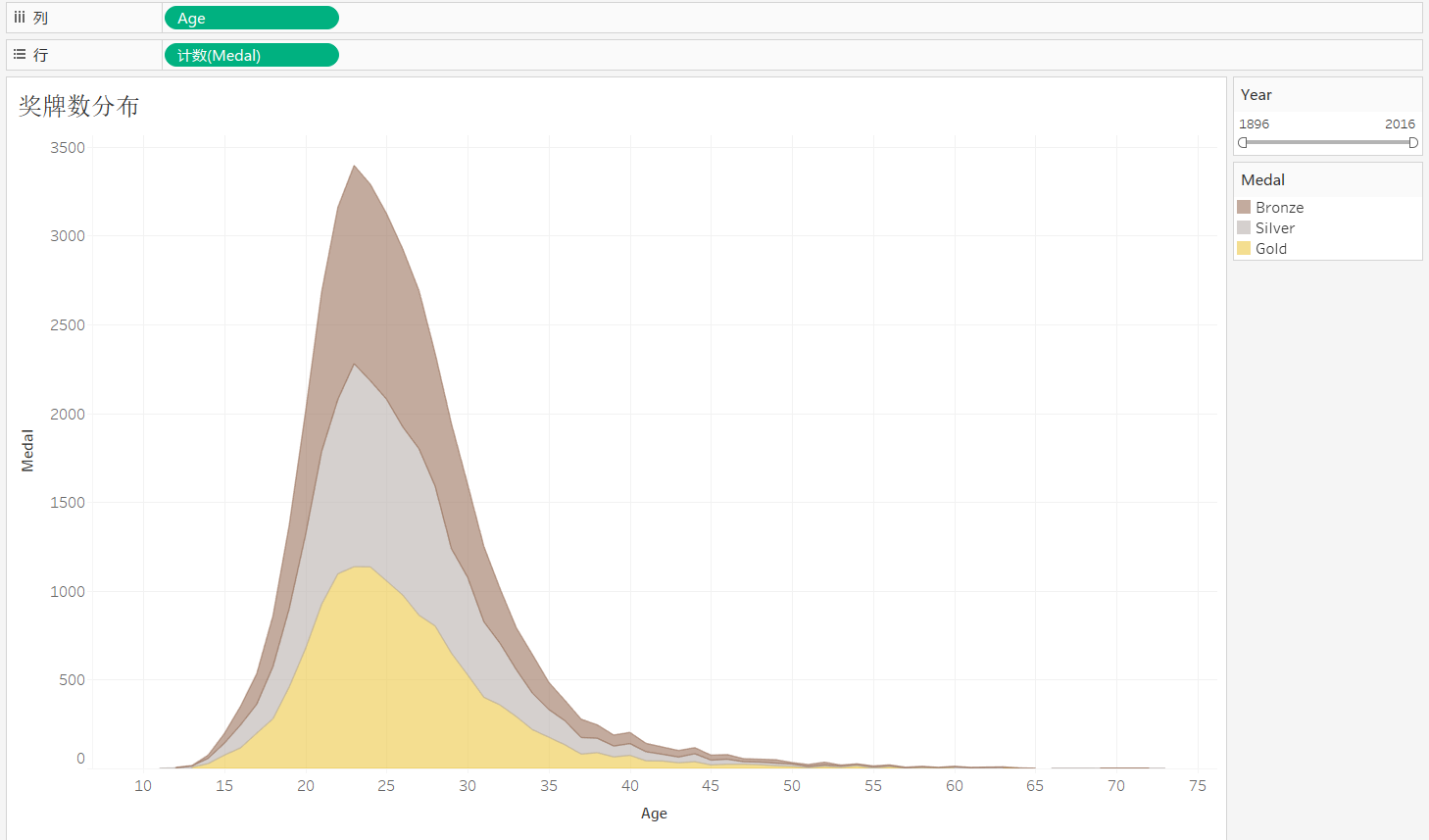


图3.1-1

(4).

绘制奖牌数排名前20国家的得奖情况的柱状图。首先将Medal和Region分别拖入行列功能区，Medal类型改为度量→计数。因为国家数太多，生成的柱状图太长，图表不便于查看，下面基于新建计算字段对显示的国家进行筛选。

因为预处理对Medal的空缺值进行了填充，若直接对其进行计数，实际上结果等于“各国参与运动员数”，与期待的“各国奖牌数”不符。故先基于Medal新建一个计算字段Medal\_Num。在数据区域Medal上右击创建→计算字段，输入表达式如下图4.1-1，然后对Medal计数即可。

图4.1-1

这里只筛选显示奖牌数前20的国家。将Region拖入筛选器，自动弹出筛选器对话框，选择顶部。按Medal\_Num字段，计数，筛选出前20个国家。然后再将国家基于Medal\_Num字段降序排序。

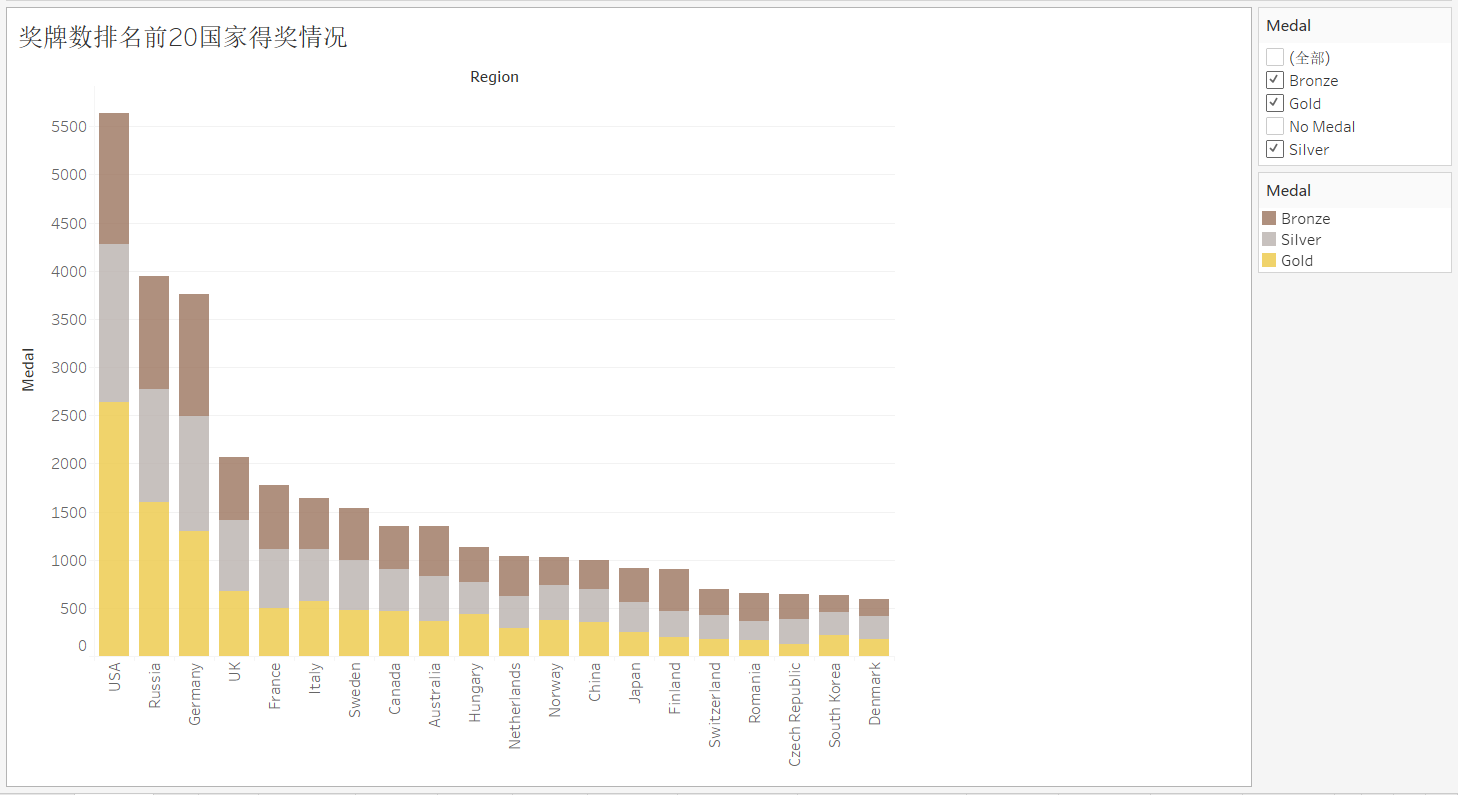
将Medal拖入标记卡中的颜色，即可生成堆叠柱状图，其中金银铜及未获奖按不同颜色显示。再将Medal拖入筛选器，选择不显示"No Medal"；在筛选器上右键，选择显示筛选器，筛选器就在视图左侧显示。堆叠柱状图效果如下图4.1-2：

图4.1-2

(5).

绘制各界奥运颁发的奖牌数叠状图，将Year和Medal分别拖到行和列中，然后将Medal拖到颜色和筛选器中，筛选出金银铜三种金牌颜色，即可显示如下图5.1-1结果：

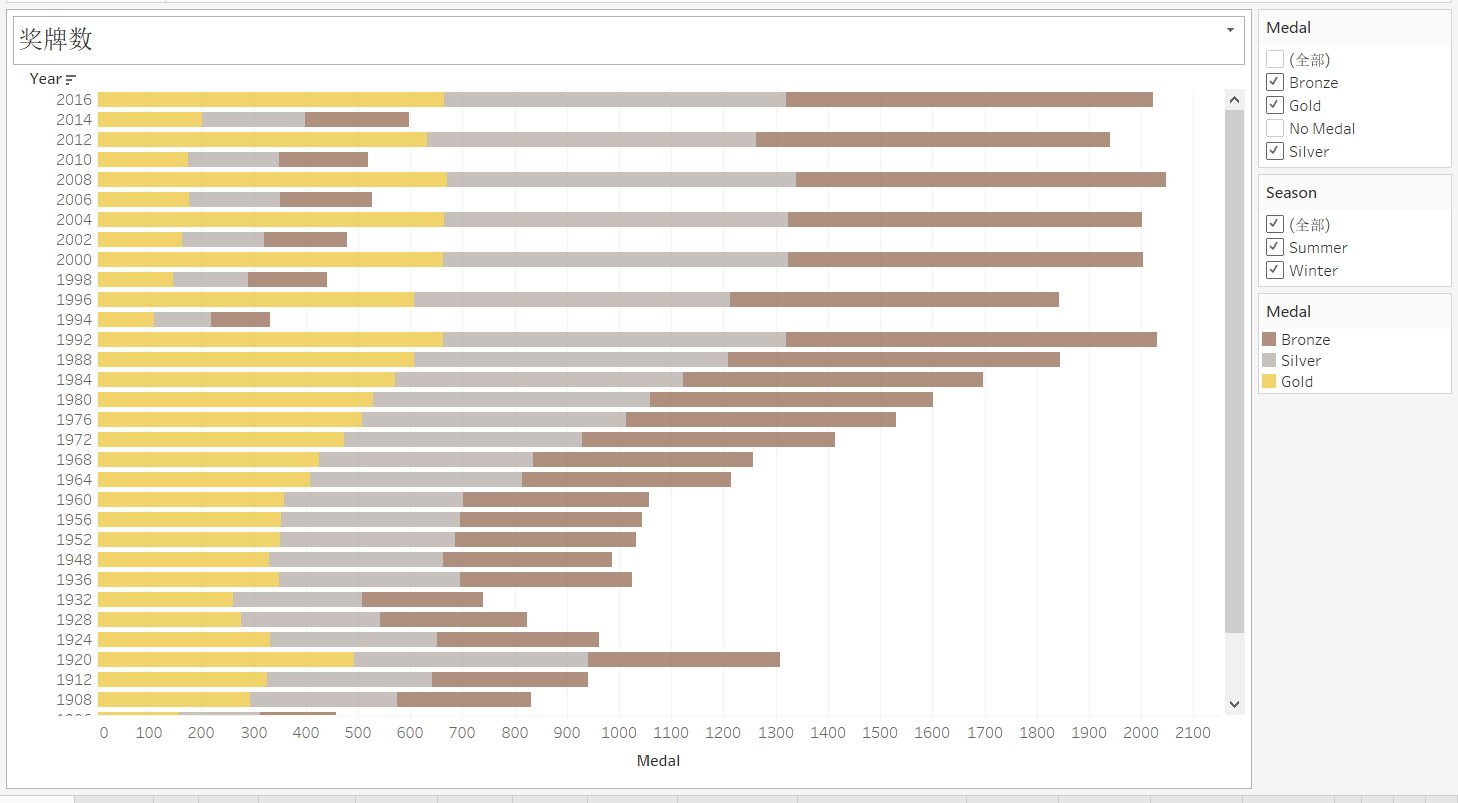


图5.1-1

(6).

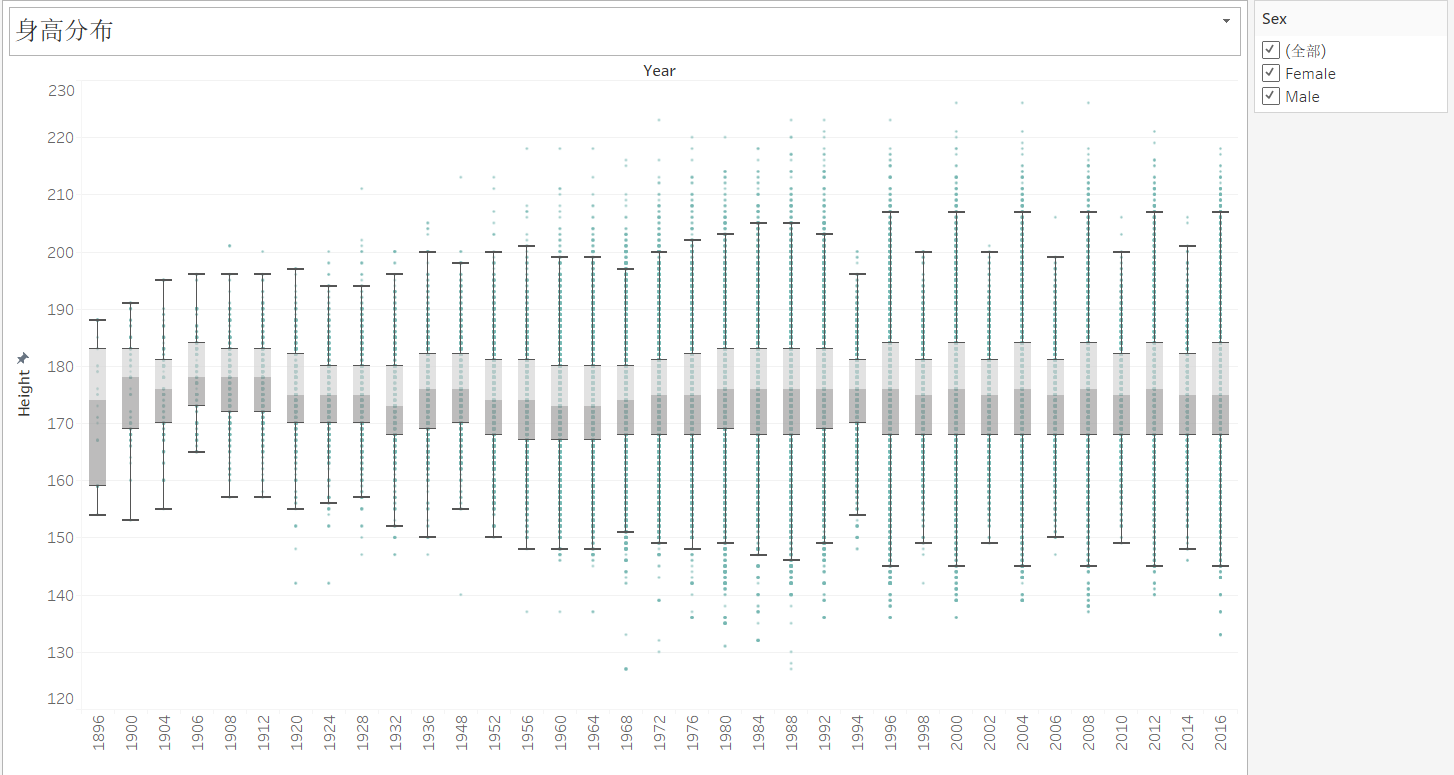
绘制各届奥运会运动员身高分布的箱线图。将Year和Height分布拖入行列功能区，在视图右侧智能推荐选择箱线图，会得到一些离散分布的线。需要在菜单分析栏，取消聚合度量，即可生成一幅箱线图。观察到数据集中于图表上部，对纵轴的坐标范围进行设置，还可添加基于性别的筛选器，最终效果如下图6.1-1所示：

图6.1-1

(7).

绘制各届奥运会运动员体重分布的箱线图。将Year和Weight分布拖入行列功能区，在视图右侧智能推荐选择箱线图，会得到一些离散分布的线。需要在菜单分析栏，取消聚合度量，即可生成一幅箱线图。观察到数据集中于图表上部，对纵轴的坐标范围进行设置，还可添加基于性别的筛选器，最终效果如下图7.1-1所示：

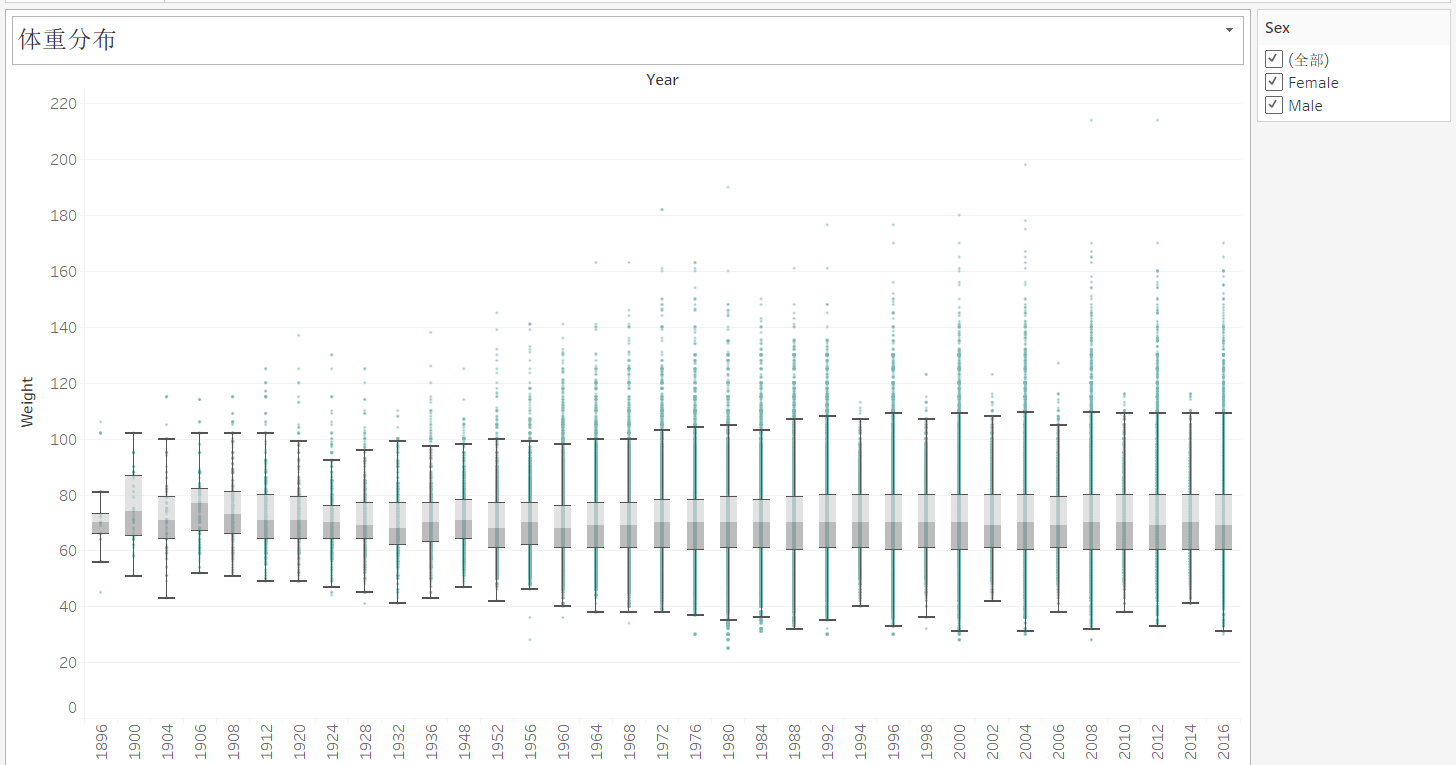


图7.1-1

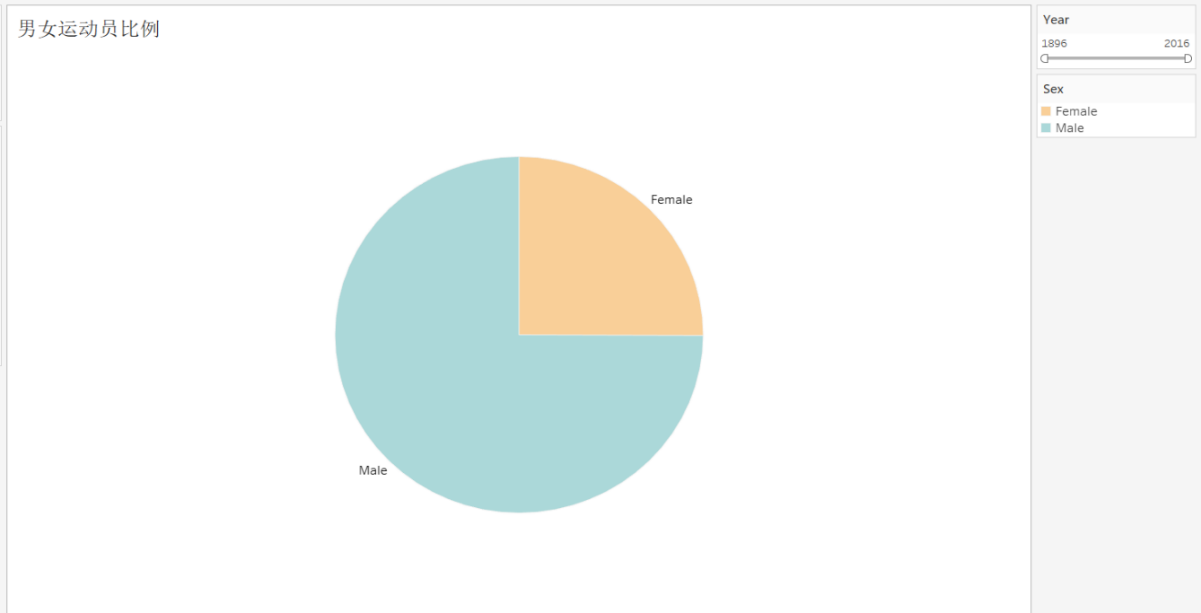
(8). 绘制男女运动员比例饼图。在标记卡中选择饼图，将字段ID拖入角度，数据类型改为度量→计数(不同)，会生成一个圆形，但是还未显示男女比。继续将Sex拖入颜色和标签，添加Year筛选器，效果如下图8.1-1所示：

图8.1-1

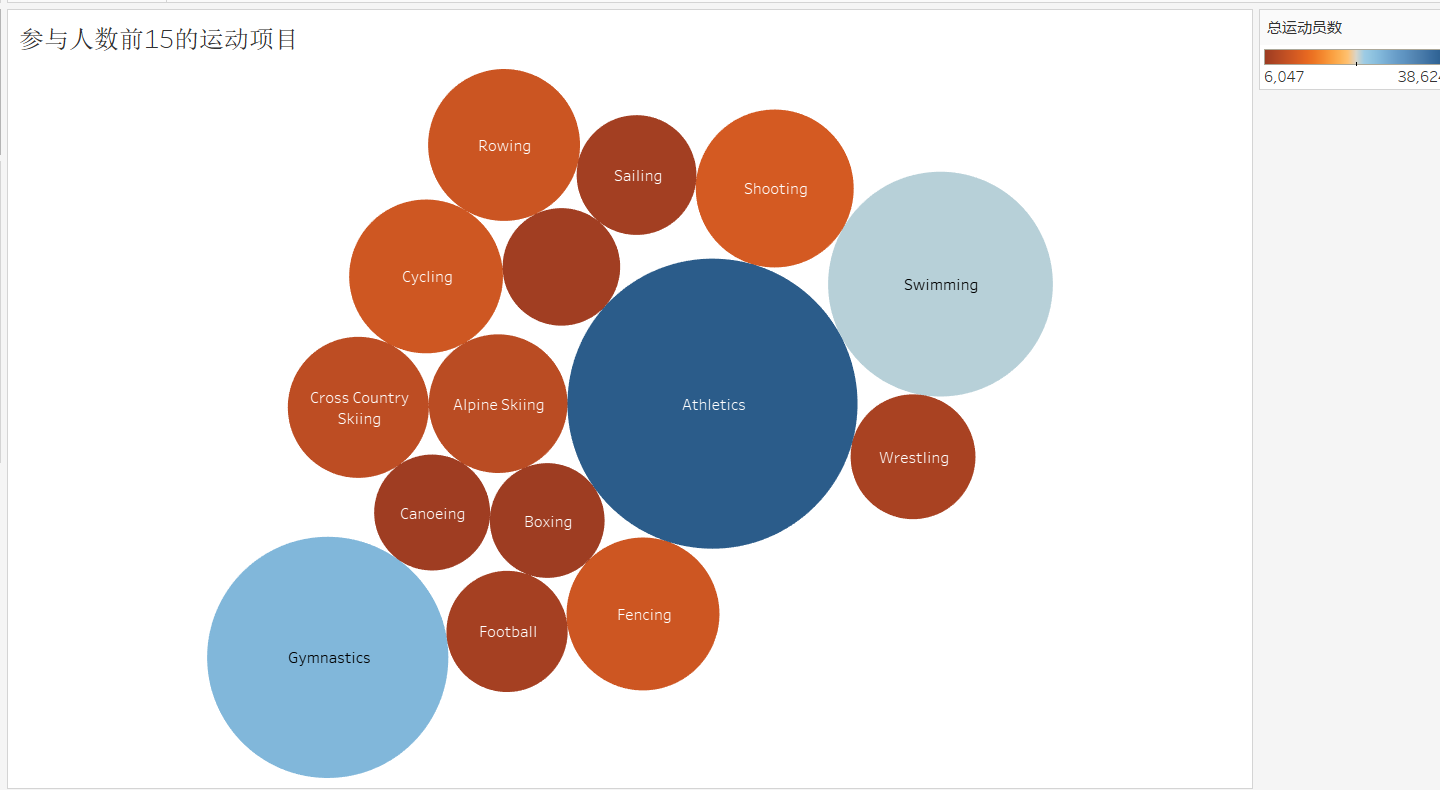
(9). 参与人数前15的运动项目，将Season，Year，Sport拖到筛选器中，设置筛选条件，然后将Sport拖到标记卡中的大小，接着设置Sprot度量为计数（不同），最后将Sprot拖到颜色标记卡中，即可显示如下图结果9.1-1：

图9.1-1

(10)．

绘制各届男女运动员人数，将Year和ID分别拖到列和行中，ID度量设置为计数(不同)，然后将Sex拖到标记卡中，最后将Year和Season拖到筛选器中，即可显示下图效果10.1-2：

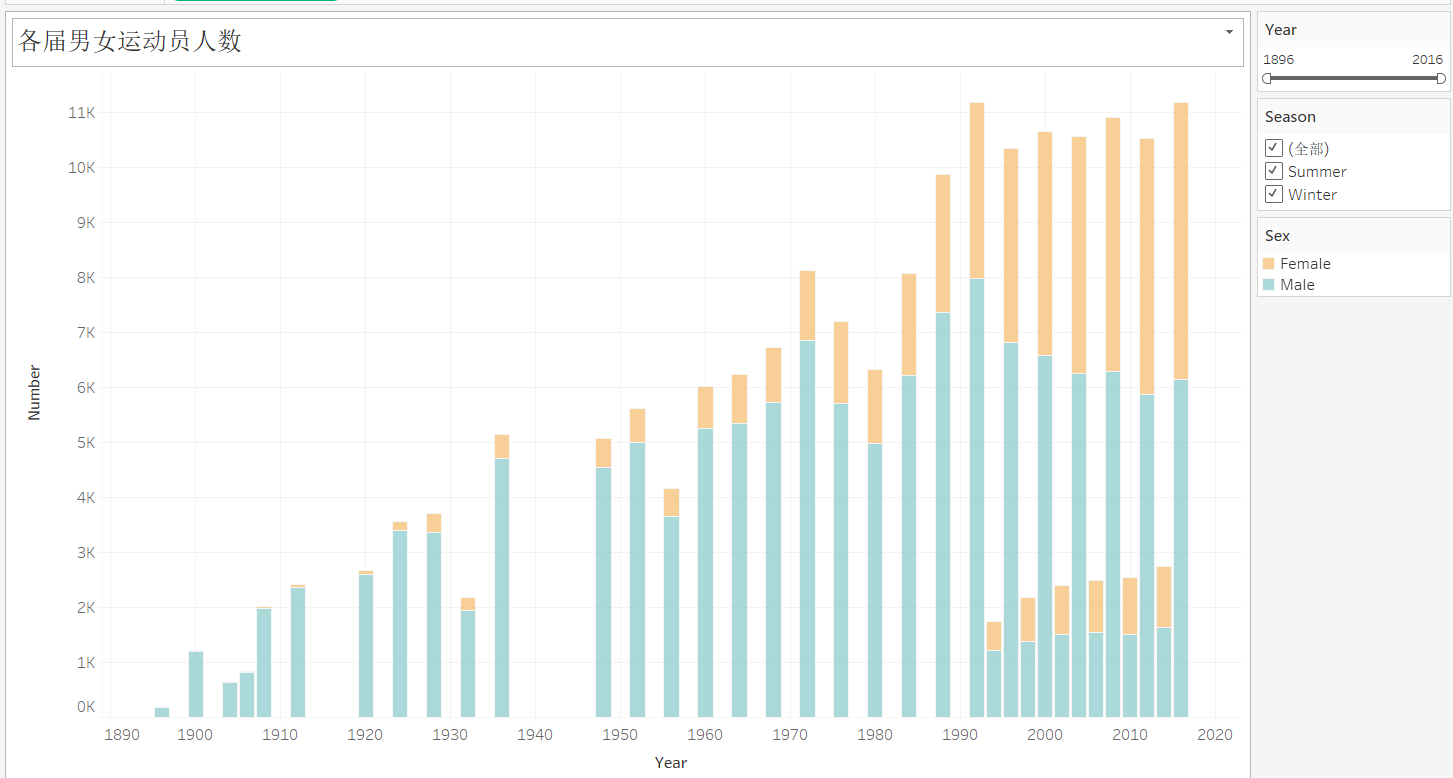


图10.1-2

(11).

绘制中国各届奥运奖牌数堆叠图，首先将Year和Medal分别拖到列和行中，并设置Medal度量为计数，然后将Medal拖到标记卡中的颜色处，最后将Region拖到筛选器中，设置筛选条件为中国，效果图如下11.1-1：

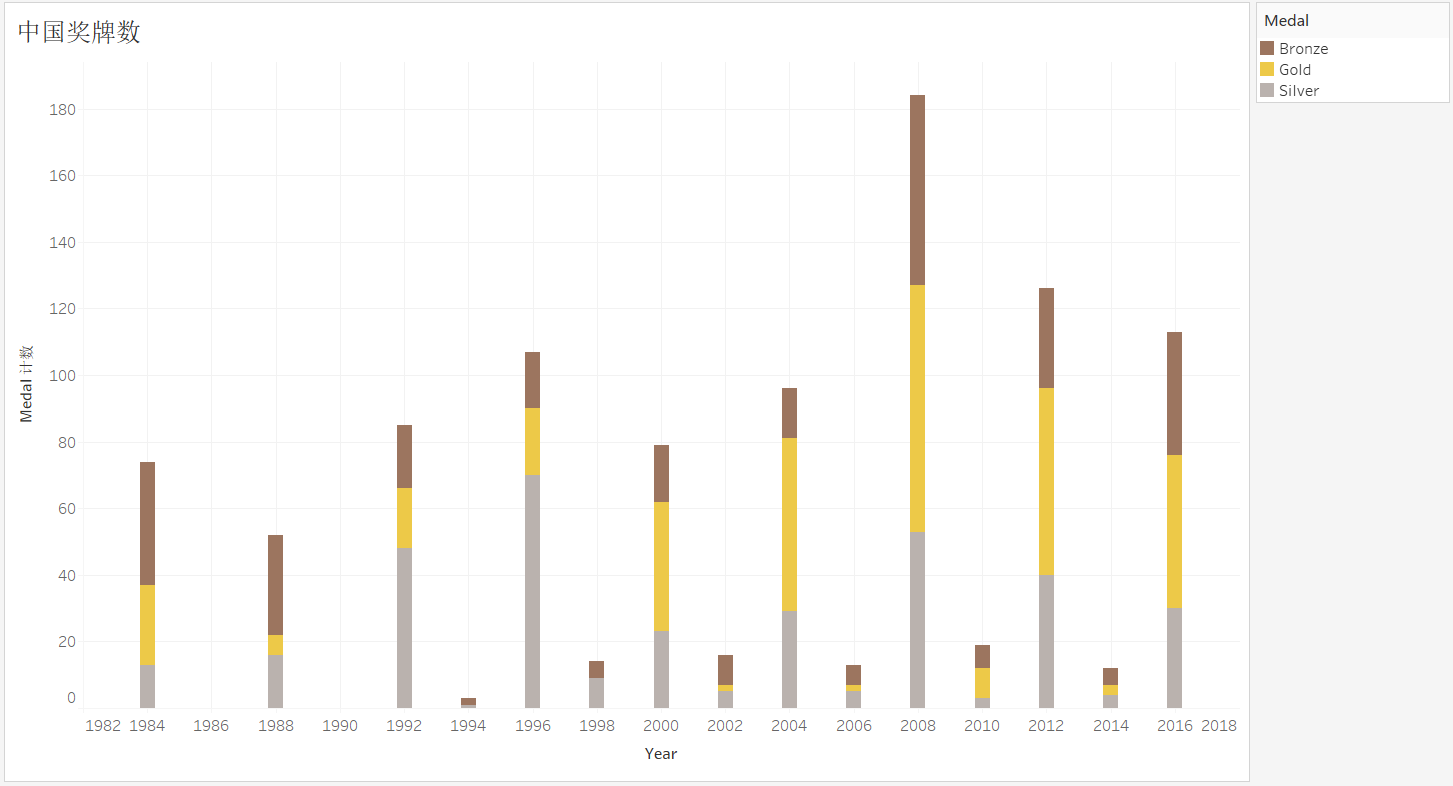


图11.1-1

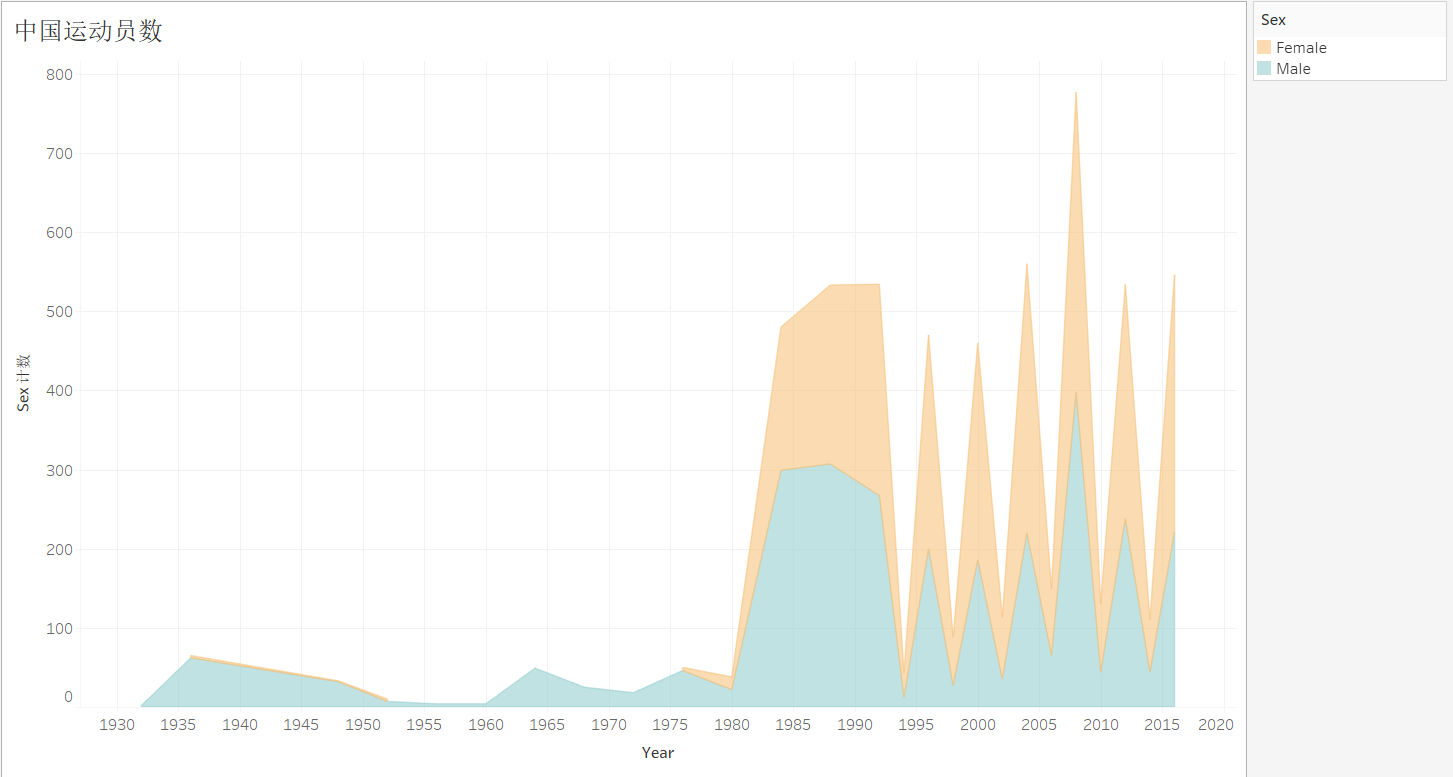
(12)．绘制中国男女运动员参加数的面积图，首先将Year和Sex拖到列和行中，接着将Region拖到筛选器，设置筛选条件为中国，然后将Sex度量设置为计数，最后将Sex拖到标记卡中，即可显示下图效果12.1-1：

图12.1-1

(13).绘制中国前二十冠军项目条形图，首先将Event和Medal分别拖到列和行中，并设置Medal度量为计数，接着将Region拖到筛选器，设置筛选条件为中国，然后将Event拖到筛选器中，筛选前二十项夺冠项目，最后并将计算字段MedalNum拖到筛选器，显示效果如下图13.1-1：

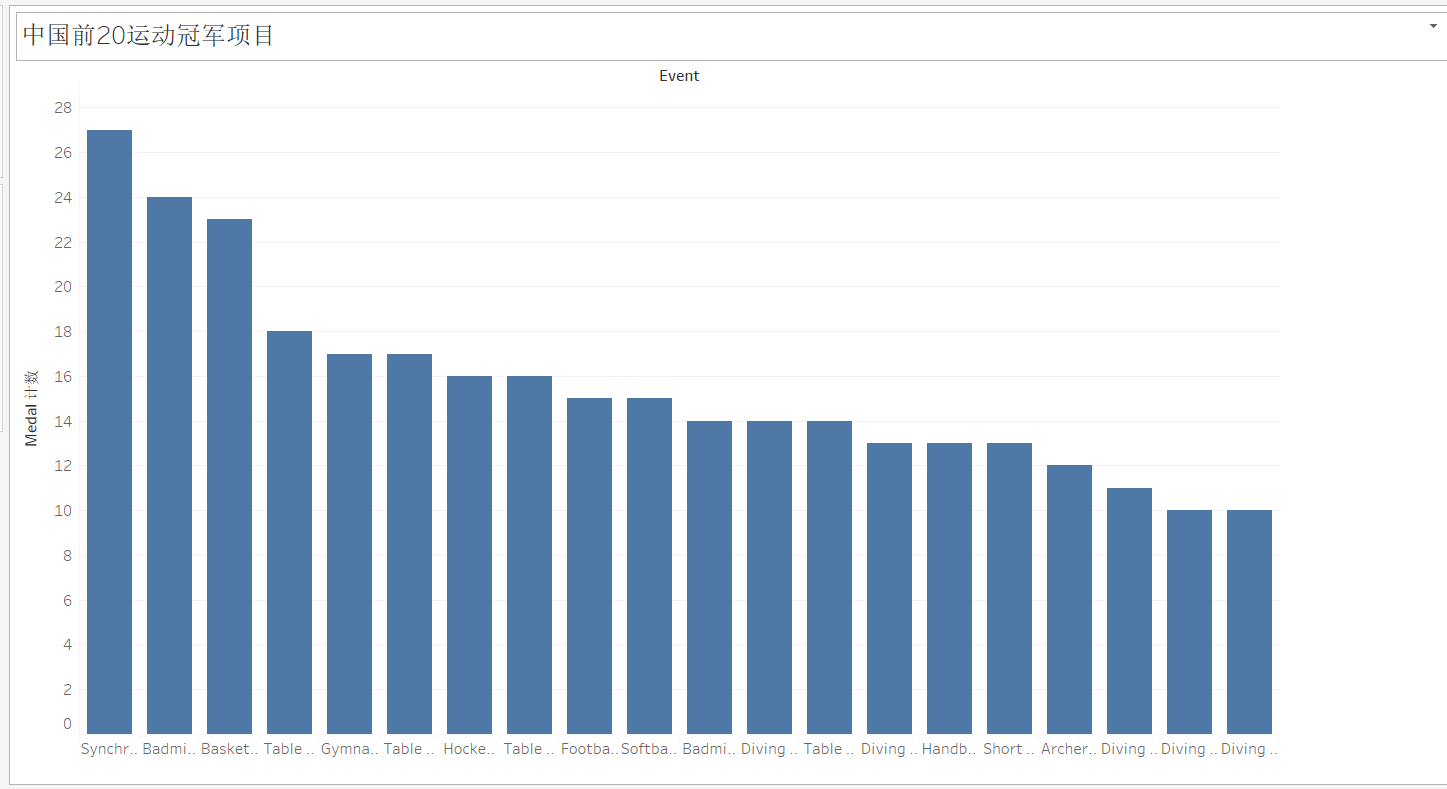
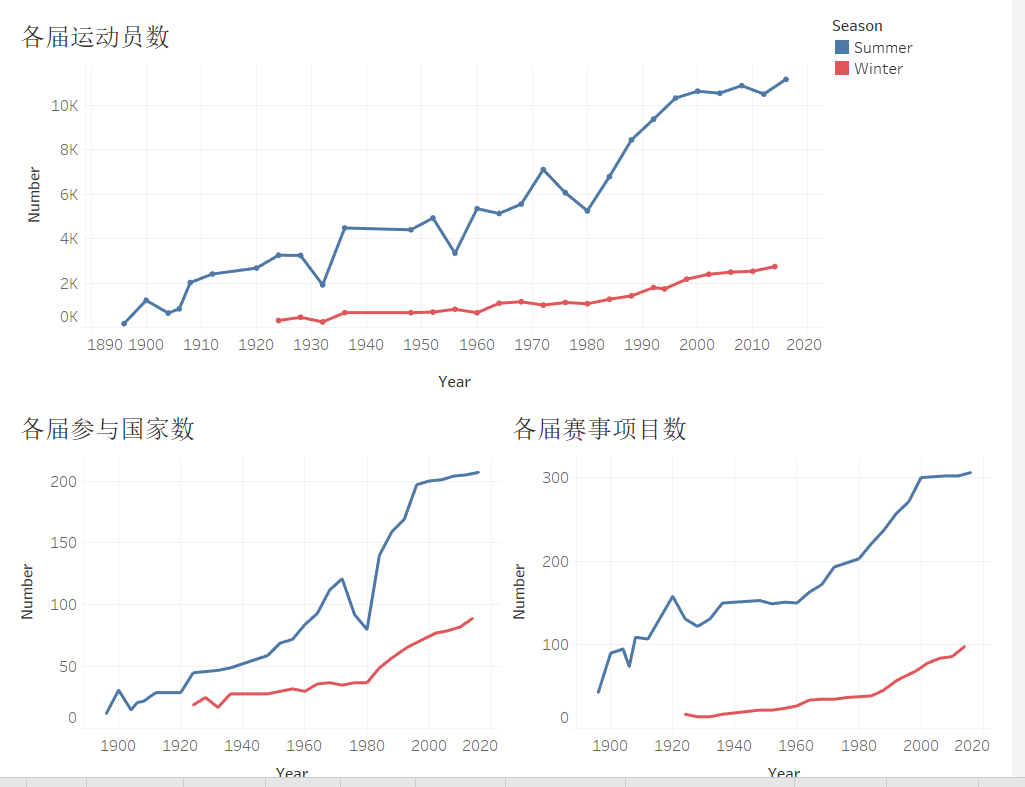


图13.1-1

**（2）我建立的仪表板是**

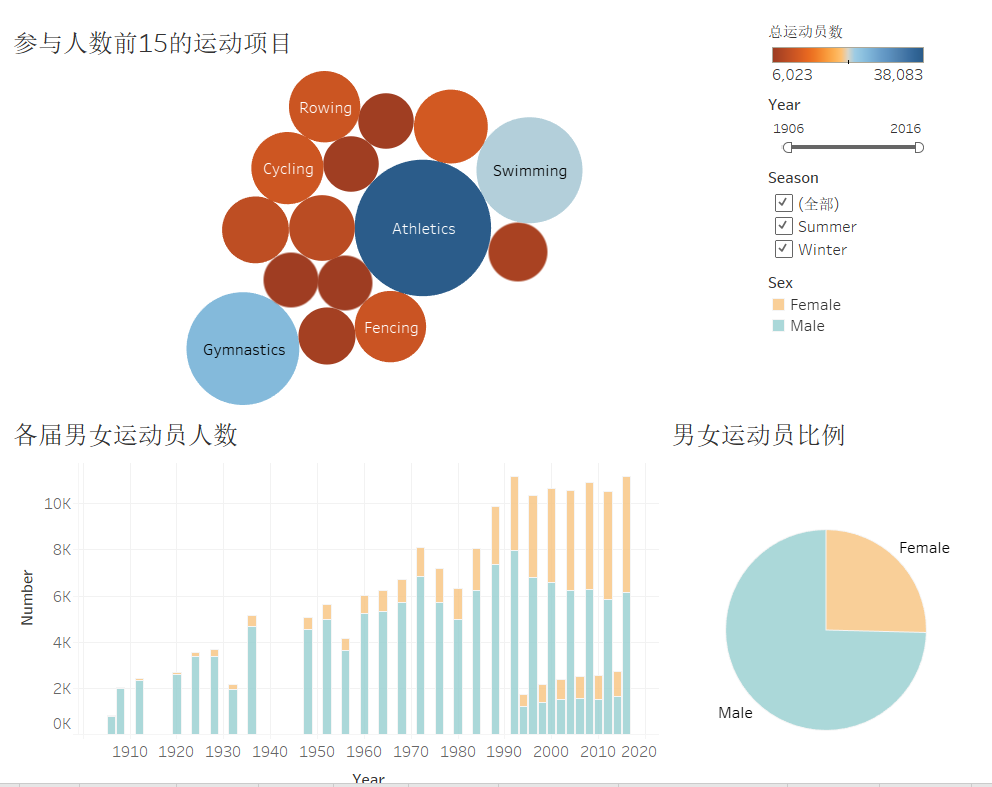
仪表板1：总体概况



图（2）.1-1

这个仪表板说明了随着时间的推移，各届奥运会的参加人数逐渐增加，奥运会重视程度在各国面前引起重视。

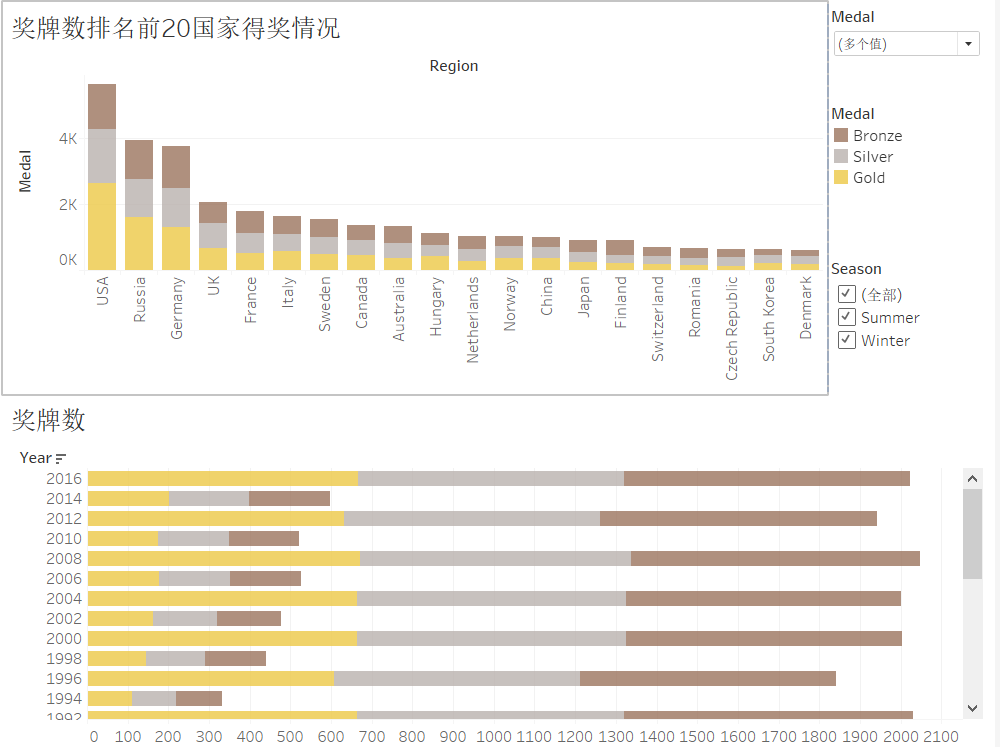
仪表板2：各项目运动员分布情况



图(2).2-1

这个仪表板说明了各届奥运会参与的男性比例远大于女性，其中大多数运动员参与田径运动和，体操运动，和游泳运动。

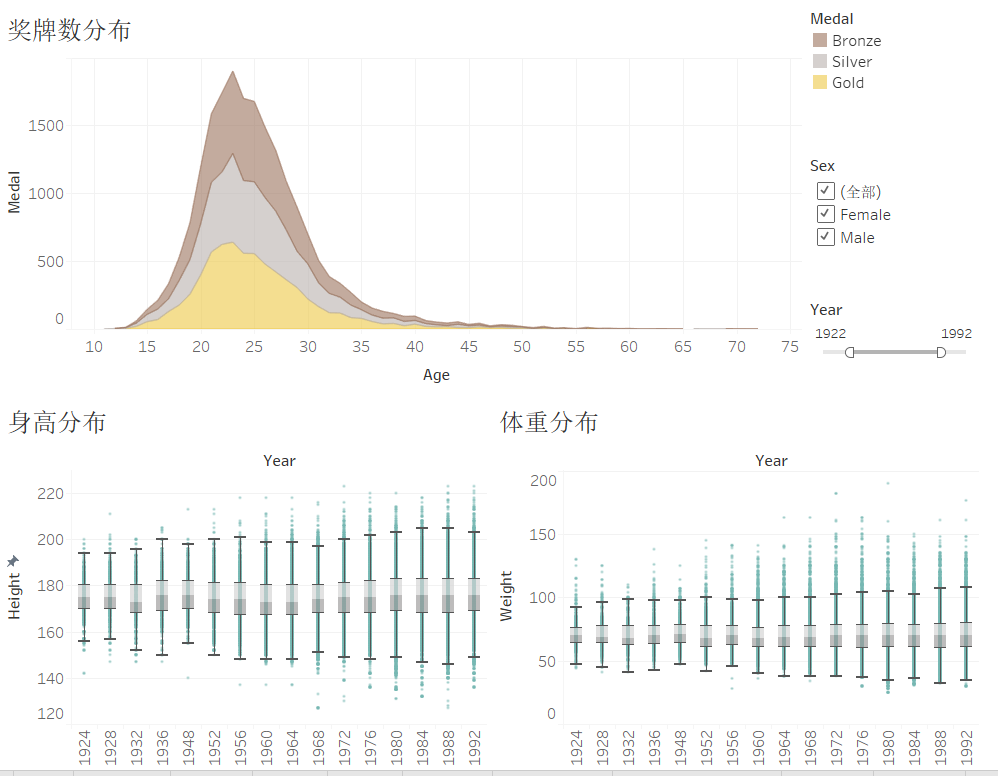
仪表板3：奖牌数在前二十的情况名国家的情况



图(2).3-1

这个仪表板说明了在这个120的奥运发展史里，获得奖牌的运动员人数越来越多，奥运会逐渐受到各国的重视，其中美国奖牌数量雄踞榜一。

仪表板4：运动员身体状况

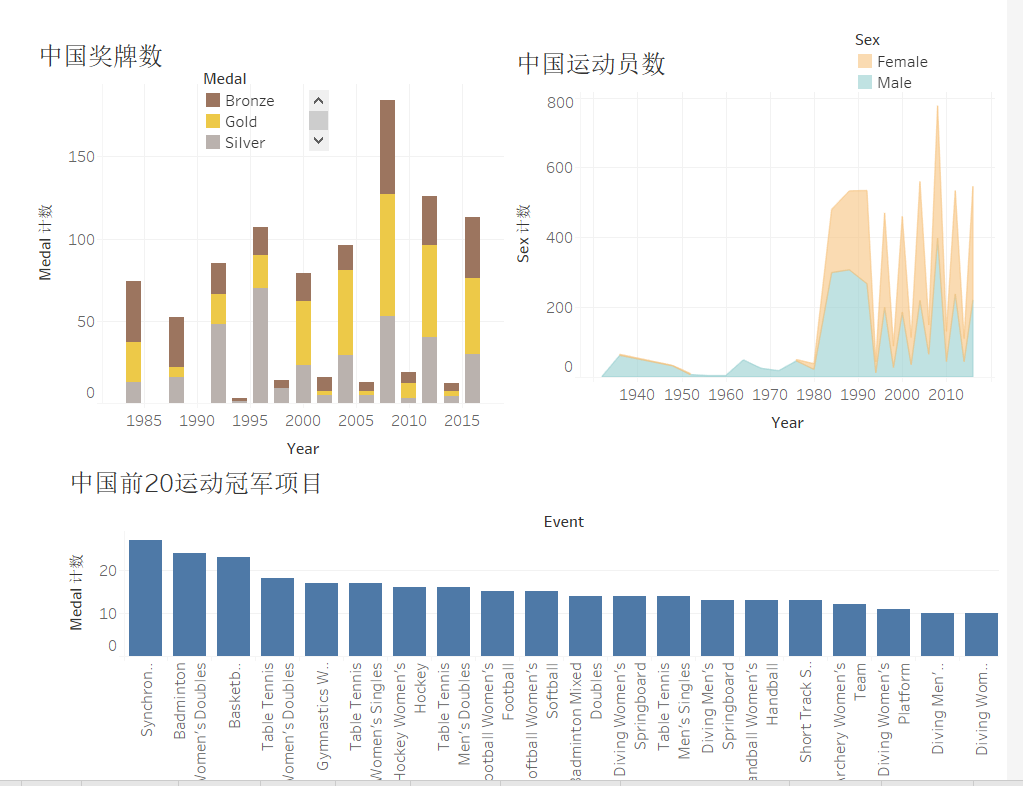


图(2).4-1

这个仪表板说了奥运获奖人数的年龄一般在25岁左右，其中他们的身高平均在170左右，体重在70kg左右。

仪表板5：中国奥运情况

这个仪表板说明了，中国逐渐重视奥运会，发展运动员，从初期只有男性参加奥运会到男女参与奥运会比例几乎相等，其中中国主要在Synchronized Swimming Women's Team（女子花样游泳），Badminton Women's Doubles（女子羽毛球双打），乒乓球等项目中夺得金牌。

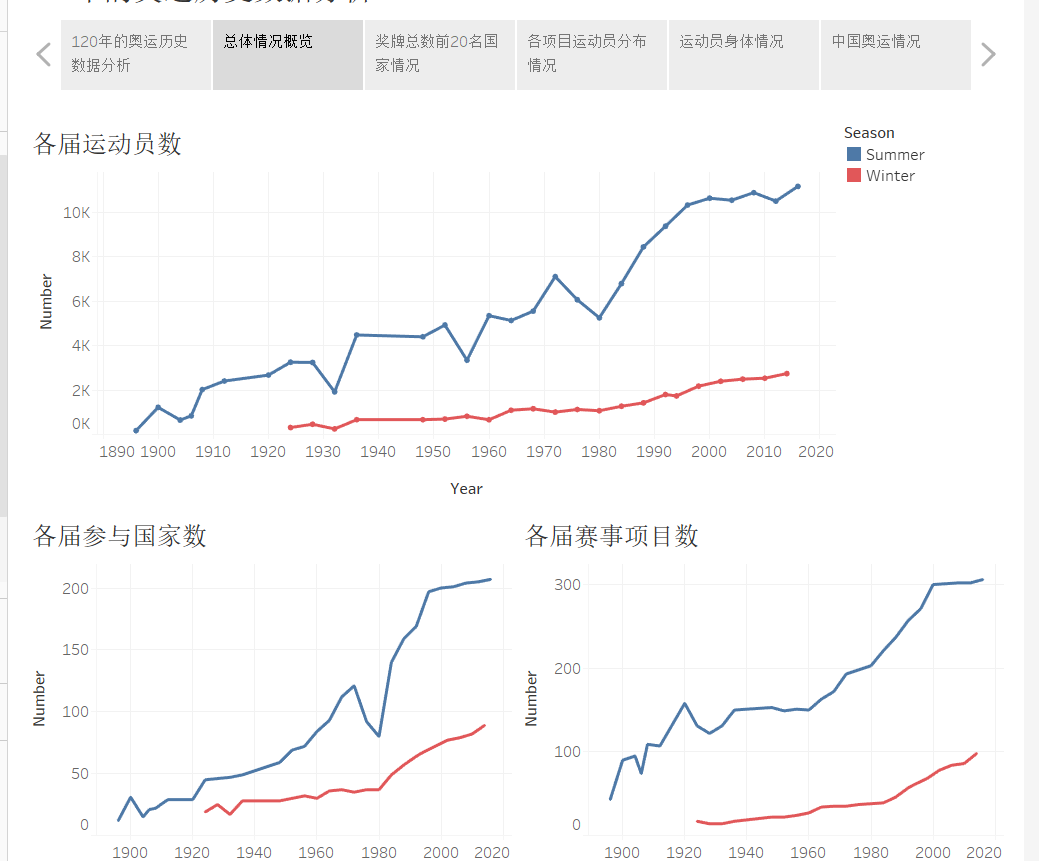


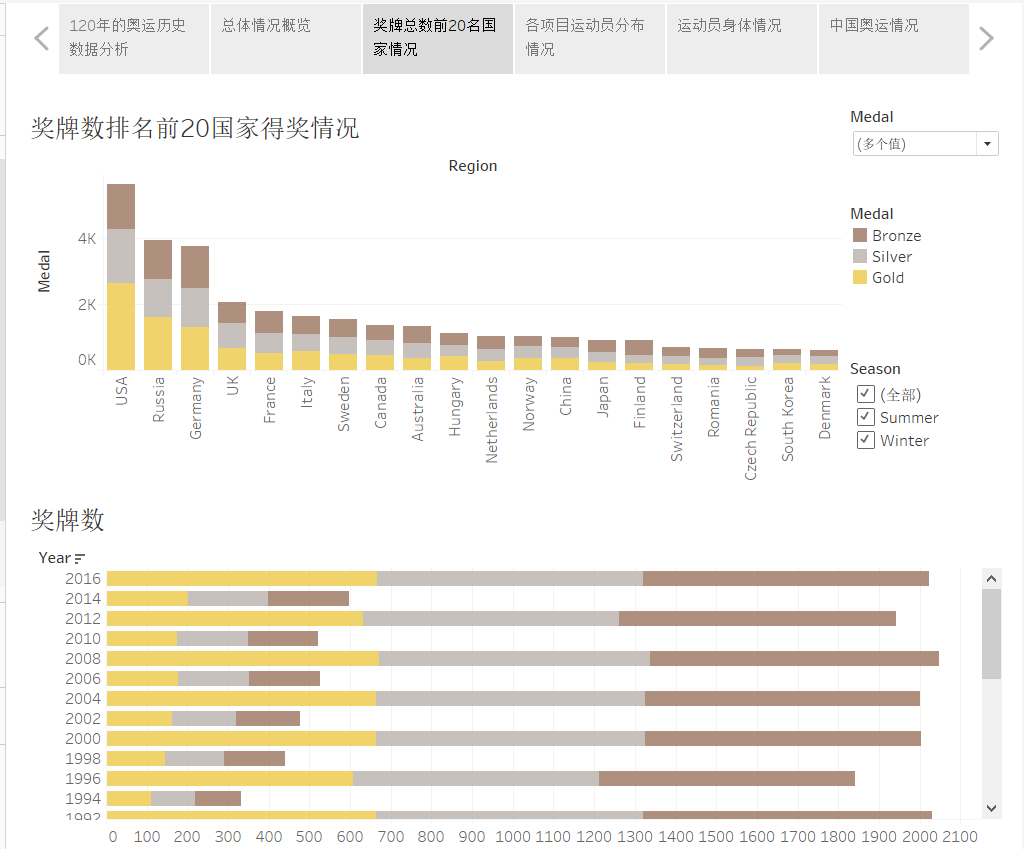
图(2).5-1

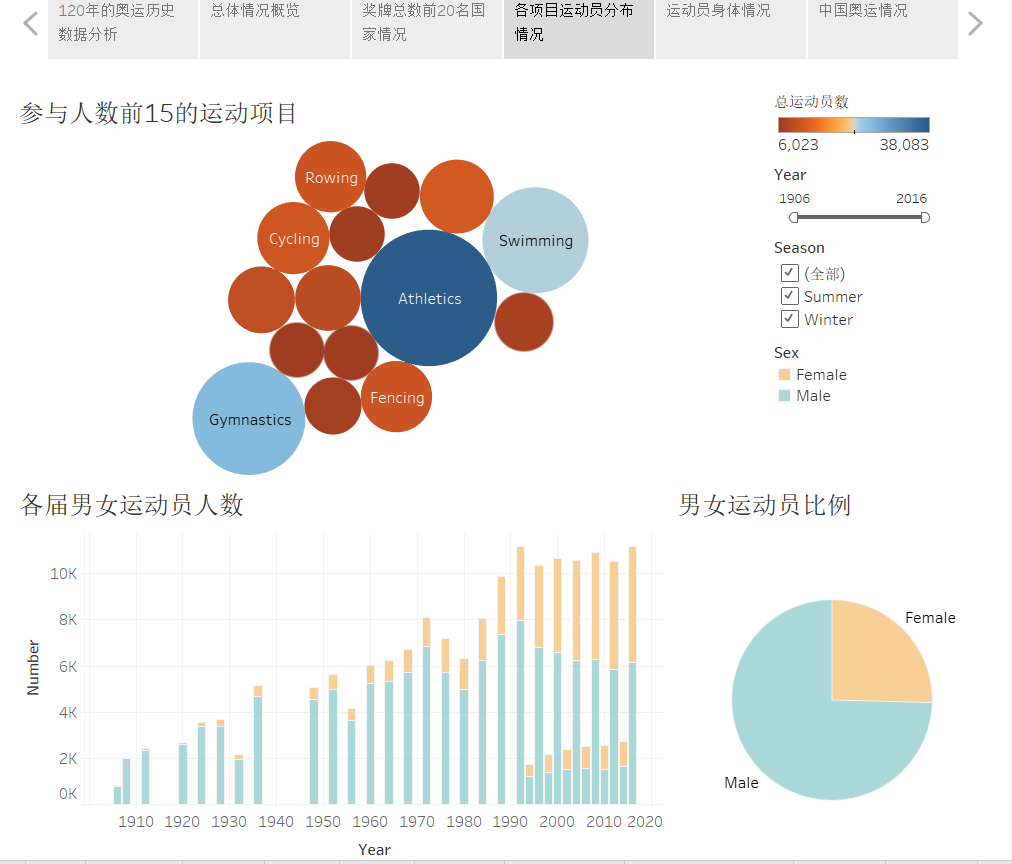
**（3）我建立的故事：**

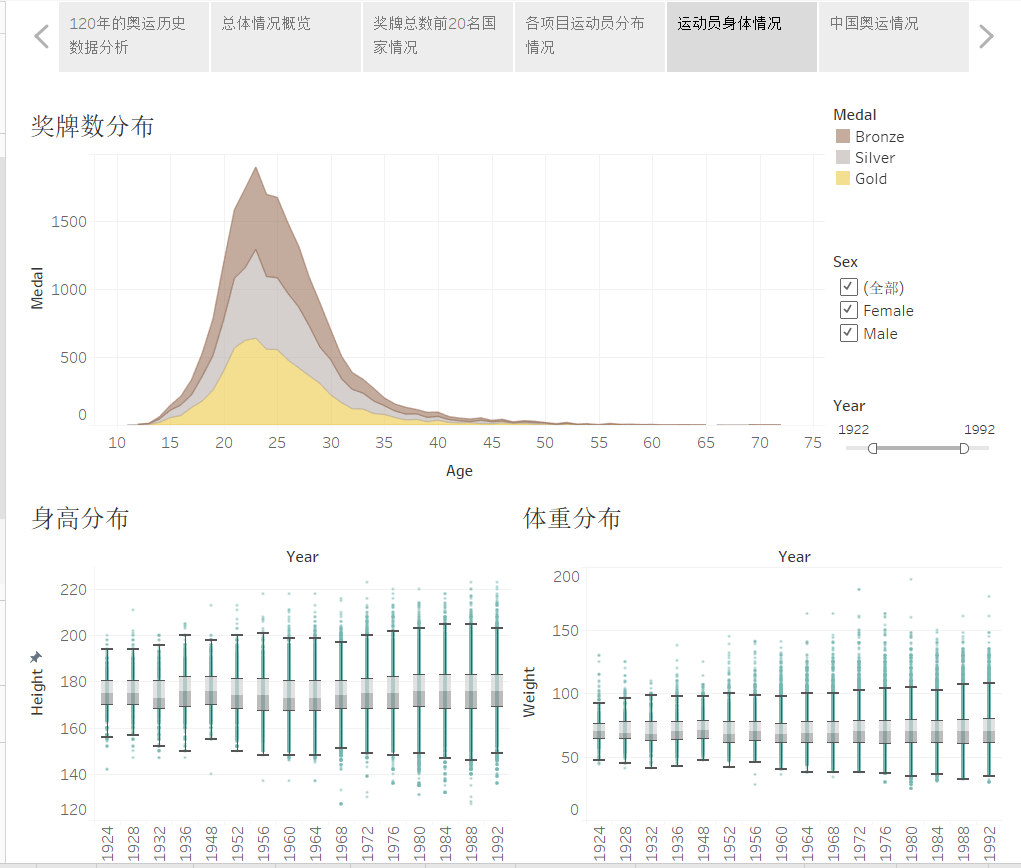


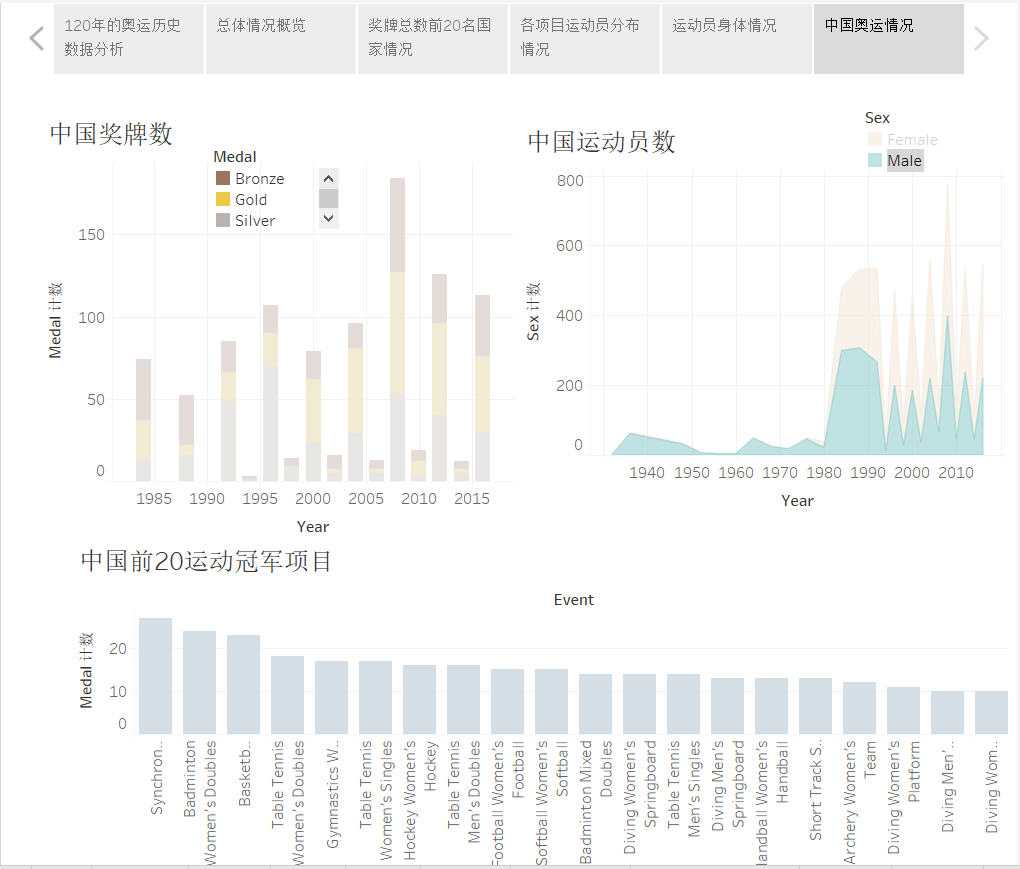
图(2).1-1











这个故事讲述了世界120年奥运史的总体概况：奖牌前二十的国家和运动员的分布情况，以及运动员的身体健康情况和中国奥运的发展情况。体现出了各国逐渐重视奥运会，发展奥运精神。

**（4）通过对120的奥运发展史的可视化分析，我获得的数据发现（信息）有：**

1. 中国队1984年开始有奖牌记录，奖牌总数逐年上升，在2008年主场举办的奥运会上奖牌数达到最高，此后一直稳定在奖牌榜和金牌榜前列；

2. 从我国收获金牌数量最多的运动项目来看，排在前三的分别是：跳水、乒乓球与体操，从排名前十的运动项目来看，大多数都是具有传统优势的项目；

3. 如果单论集体项目（足球、篮球、排球、手球、曲棍球）的金牌数量排行，无疑是非常能展示大国的体育实力的，这也是衡量一个国家是否体育强国的硬性指标。从该项指标来看，美国383枚金牌遥遥领先，而我国仅有34枚，排名第18，而且全部为中国女排获得，其他项目则一无所获，说明了我国拿金牌的实力虽然名列前茅，但在热门的集体运动项目里却严重缺乏竞争力；

4.随着时间推移，奥运会赛事数量整体呈现上升的趋势，且夏季奥运会赛事数量明显远超与冬季奥运会赛事数量；1896年开始到1980年夏运会参赛运动员数量呈现曲折上升的趋势，这是因为此时处于世界格局动荡期，战争、国家间冷战摩擦等事件很大程度影响了奥运会举办的客观条件，同理冬运会也类似；

5. 男性夏运会参赛运动员人数是和历年运动员数量变化同调的，女性夏运会参赛运动员人数明显的持续增长，并在1980年到2000年间持续大幅上涨。120年来的奖牌数量排名如下，USA为最多，CHN（中国）为第17名。