|  |  |
| --- | --- |
| **评分标准** | **评分** |
| **（1）真实性和可用性：数据是真实可用的、具有可用性；** |  |
| **（2）感知性和可行性：易于感知，方案可行；** |  |
| **（3）艺术性和设计性：布局合理、可交互、美观展示；** |  |
| **（4）原创性和有价值：实用性、内容新颖详实；** |  |
| **（5）论文内容合理、书写工整。** |  |
| **总分** |  |

# 电子游戏销售的可视化与交互

**学号：19121525 姓名：蒋俊杰**

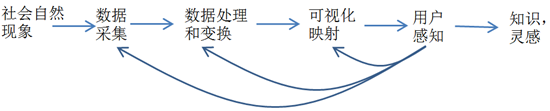
**摘要：** 在21世纪的当下，电子游戏在人们生活中的占比越来越大，本文用于可视化的数据是利用爬虫软件从vgchart.com网站上得到的，主要描述了全球市场上电子游戏的销售情况。

**关键词：**电子游戏,plotly,echarts

# 引言

电子游戏（Video Games，少部分学者使用Electronic Games）又称电玩游戏（简称电玩），是指所有依托于电子设备平台而运行的交互游戏。根据媒介的不同多分为五种：主机游戏（或称家用机游戏、电视游戏）、掌机游戏、电脑游戏、街机游戏和移动游戏（主要是手机游戏）。一般市场上我们所接触的游戏都是电子游戏。或许你一直在接触各种不同类型的电子游戏，但却从来没有思考过我们喜欢玩的游戏它的销量和质量都是由哪些因素决定的。在本文中我将从出版商、发售平台、地区销量等各个方面通过数据可视化的形式来分析电子游戏的销售情况。

# 电子游戏销售可视化的理论设计和数据采集



在经过一学期的数据可视化与人机交互课程的学习之后，在设计一个数据可视化之前，我首先明确的一点是不能盲目的、想当然的构思一个数据可视化的方案，而是需要在课程理论的学习基础上，通过一定理论的支持来有逻辑的设计一个数据可视化的方案。因此本文在《数据可视化的基本原理与方法》一书中的第一章、第三章、第四章以及第五章第五节的理论基础上进行了如下数据可视化的设计。可视化流程如上图所示。本部分主要介绍本文的可视化设计和数据来源。

2.1源代码存放地址为：

<https://github.com/JJJtriple/->

2.2社会自然现象的观察：

首先我们要明确：数据可视化的目的是洞悉蕴含在数据中的现象和规律，这包含多重含义：发现、决策、解释、分析探索和学习。因此数据可视化的全体既要在数据上有价值性，又不能背离了社会现象。本文的选题是电子游戏销售的可视化，因为生活在21世纪，人们可能或多或少的都有接触过电子游戏的相关经历，但对游戏的销售趋势以及受欢迎程度背后的原因却没有深入思考过。或许对于普通玩家们这是不需要在意的事情，但对于追求销售额的游戏出版商或者是购买相关游戏公司股票的股民们来说，这些问题就显得较为重要了，甚至优秀的游戏公司中可能有专人负责相关方面的分析。本文就将从电子游戏的类型、发行方式、销售地区等因素方面从时间线上来研究销售额的变化并做出区别于单纯的数据形式的可视化报告。

2.3数据采集：

数据是可视化的对象，数据可以通过多种多样的方法来采集。在明确了任务需求之后，我们的数据采集才能在数据格式、维度、尺寸等因素上符合要求。因为这些因素很大程度上决定了可视化结果的质量。本文的数据采集方式是通过爬虫程序来进行的。相关的爬虫代码已经上传到了前文的提到的连接中。数据的来源是<https://www.vgchartz.com/>，这是一个全球范围内较为权威的电子游戏销售数据统计网站。根据网站显示的信息，本文选择采集的数据在字段上包含了：RANK-总销售额的排名、Name-游戏的名字、Platform-游戏发布平台(即PC,PS4等)、Year-游戏发行的年份、Genre-游戏的类型、Publisher-游戏的出版者、NA\_Sales -北美销售额(百万)、EU\_Sales -欧洲销售额(百万)、JP\_Sales -日本销售额(百万)、Other\_Sales—世界其他地区销售额(百万)、Global\_Sales-全球销售总额。

# 电子游戏销售数据集和可视化的展示

在前文介绍了数据来源之后，在本部分将展示数据处理和变化的过程、可视化的映射以及用户的感知。

3.1数据处理和变换：

数据的处理与变化可以处理原始数据不可避免的噪声和误差，为后续的可视化映射做前期的准备工作。本文的数据处理与变化在爬虫软件的编写中就开始进行，因为浏览网页时可以发现有些字段有着较多的缺失值，因此必须采取清洗操作，例如对日本销售额的抓取：

|  |
| --- |
| sales\_jp.append(  float(data[11].string[:-1]) if  not data[11].string.startswith("N/A") else np.nan) |

在得到爬取的.csv文件后选择使用python打开，使用其中的numpy和pandas库来初步对数据的模式和特征进行提取。

表格

描述已自动生成

可以看出Year-年份、Publisher-游戏的发行商两个字段有所缺失，所以必须删除缺失值。删除方式及结果如下：

图片包含 文本

描述已自动生成

3.2可视化映射与用户感知：

在上图中可以清晰的看出采集到的数据以及各个字段的内容。接下来就是可视化流程的核心部分：可视化映射的介绍。可视化映射一方面与数据的数值、空间坐标、不同数据间的连接等有关联，一方面也与之后用户直观感受的清晰与否有着直接的联系。可视化的映射的最终目的当然是让用户可以洞察数据背后隐含的现象和规律。本文采取的可视化的工具包含了python中的plotly库以及echarts工具，两者都有可视化与交互功能。而在具体的可视化展示之前，我首先提出几个疑问：随着时间的推移，全球各地的电子游戏销量如何？在全球如此多家游戏发行商中有哪些具有代表性的厂家？为什么他们具有代表性？从游戏类型的演变来看游戏类型的流行趋势？基于这些具有实际意义的问题来进行下面的可视化映射。

首先分析数据的相关性，通过销售额与其它要素的热力图来判断相关性：

图片包含 网站

描述已自动生成

由上图可见日本地区的销售额与年份呈现负相关。在分析了相关性之后，我们回到之前提到的那些问题，首先，全球各地区的销售趋势如何？想知道趋势那必然要让销售额虽年份变化，那么就必须明确数据中包含的所有年份：

图形用户界面, 文本, 应用程序

描述已自动生成

之后我们再根据年份利用echarts将销售额与地区之间的关系利用折线图与饼图展现出来：

图表

描述已自动生成

除此以外我还绘制了每年游戏发布的数量与全球游戏销售额的变化柱线结合图：

图表, 直方图

描述已自动生成

紧接着为了考虑下一个部分的问题：具有代表性的游戏发行商有哪些？为什么他们受欢迎的程度最高？我绘制了排名前十的出版商发布的游戏总量与出版商游戏销量之间的柱线结合图，以及五大发行商的历年销售变化柱状图：

图表

描述已自动生成

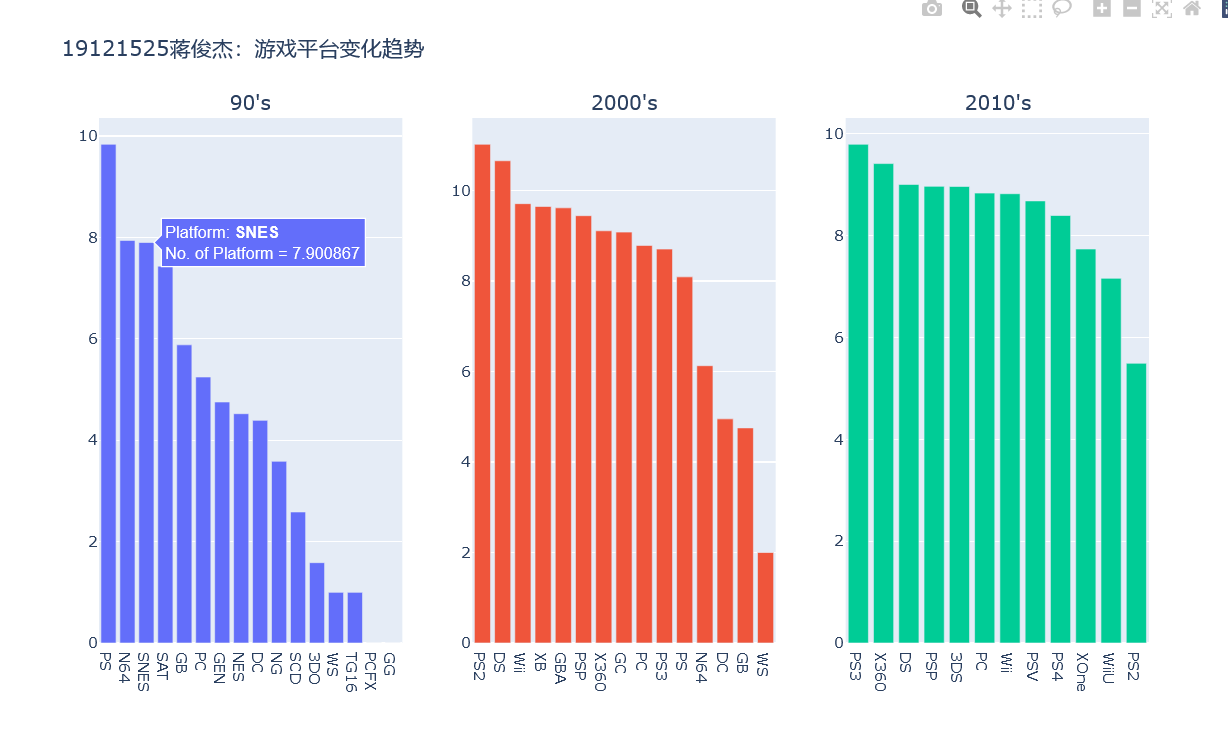
图表, 瀑布图

描述已自动生成

那么这些发行商排名存在差异的背后必然有着相关原因，或是因为推出了爆款游戏而大卖或是因为发布游戏多以量取胜，不管怎么样都离不开用户的支持与喜爱。接下来的可视化就是沿着这个思路，思考用户受欢迎的游戏类型是哪些，用户常用的游戏平台有哪些，这与出版方排名之间又存在着怎样的联系。为此我绘制了如下：游戏类型的数量随着年份变化的柱形图、游戏平台上发布的游戏数量随着年份变化的柱形图，以及游戏厂商与游戏类型之间的雷达图（较大的数值为了显示的直观性取了对数数值）：

图表, 条形图

描述已自动生成



图表, 雷达图

描述已自动生成

# 本设计的特点

# （1）真实性和可用性

本文使用的数据来自于真实的电子游戏销售统计网站：<https://www.vgchartz.com/>并且利用爬虫程序，依据有实际意义的问题，选定了科学的数据维度，数据字段的含义清晰，并且在后续的可视化与用户感知中起到重要的作用：RANK-总销售额的排名、Name-游戏的名字、Platform-游戏发布平台(即PC,PS4等)、Year-游戏发行的年份、Genre-游戏的类型、Publisher-游戏的出版者、NA\_Sales -北美销售额(百万)、EU\_Sales -欧洲销售额(百万)、JP\_Sales -日本销售额(百万)、Other\_Sales—世界其他地区销售额(百万)、Global\_Sales-全球销售总额。并且在开始数据可视化之前，对于数据的缺失部分进行了清洗，并且在可视化中对于数值较大的数值进行了数据变化使得数据更加具有可用性。下图展示了原网页的信息：

图形用户界面

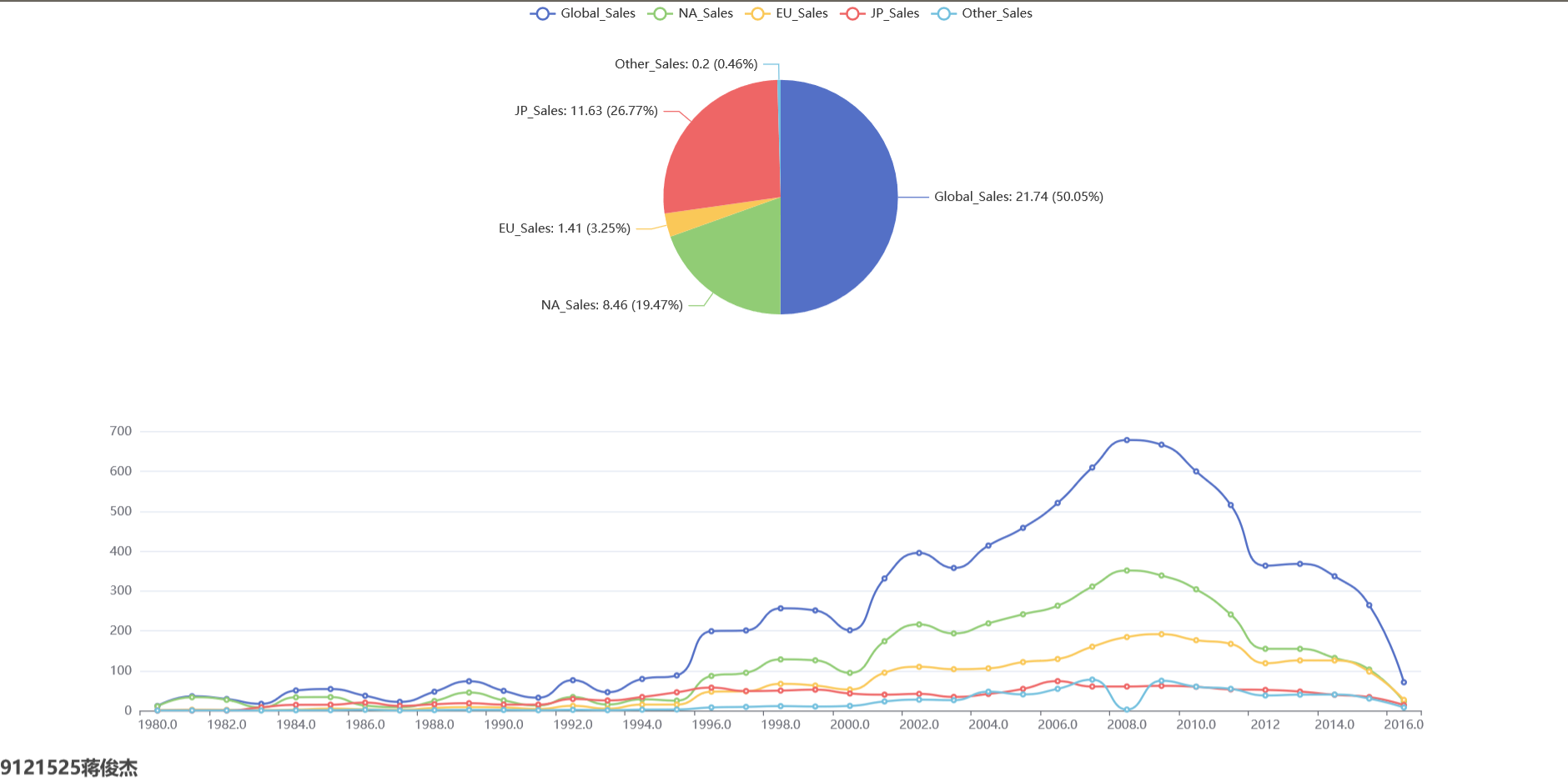
低可信度描述已自动生成

# （2）感知性和可行性

为了提高用户的感知的直观性，本文在设计之初围绕着三大问题展开：全球销售趋势、电子游戏出版商、游戏类型与发布平台。而串联这三部分的要素就是销售额也就是用户喜爱程度。为了让读者直观感受数据之间的联系我设计了热力图，为了让读者感受全球电子游戏销售趋势的变化我使用了折线+饼图的组合形式。折线直观反映数值，饼图反映占比，简单易懂。为了让读者了解出版商的游戏总发布量和销售额我选择了柱线结合图，因为这可以很好的看出销售额与发布量的关系，直观感受的是不是游戏发布量越多销售额就越大，以便读者更加明确的体会到电子游戏行业是以量取胜还是追求质量。最后为了了解用户们喜爱的游戏类型，我分类了三大时间段，以十年为单位，统计各大平台已经游戏类型的销售图，并且绘制了前五大受欢迎的厂商在前六中用户喜欢的游戏类型方面的销售额，从而可以得出结论，一个电子游戏厂商受欢迎不是没有原因的。

# （3）艺术性和设计性

可视化有着各种各样的标准，首先是表达能力强。由于用户交互是可视化不可分割的一部分，表达力也应该把用户在交互中获得的信息考虑在内。如下图所示，我将可视化与用户的交互同时放在了饼图与折线图当中，一方面可以通过页面最上方的选择按钮选择要展示在图中的字段，可以选择是否包含Global\_Sales字段，因为如果包含了Global\_Sales字段那么这部分占比会较大，无法看出各地区的销售额的占比。



除此以外在考虑游戏发行商的销售额可视化时我采用了如下动态的柱状图，这样方便观察出数据的变化规律。

图表, 瀑布图

描述已自动生成

需要强调的是我在选择绘图的数据之前就已经进行了数据变换，需要选择出排名前五的游戏发行商本文是根据销售总额来筛选的。另外为了实现可视化的简洁与美观标准，本文没有为了追求性能设计许多的菜单、滑动按钮等，而是采用一种用户易使用的方法：只需要用鼠标选择你想要了解的数据的具体信息就会做出显示，另外动态的柱状图只需要点击播放就可以体现数据的动态变化。选择雷达图来反应销售额、游戏平台、游戏厂商直接的关系也体现了用较少的图形可视化展现较多信息的有效性标准，为了美观这里还为html页面添加了背景图片让可视化显得不突兀，图形颜色的选择也较为柔和。

图表, 雷达图

描述已自动生成

# （4）原创性和有价值

任何可视化主题脱离了现实就显得华而不实，本文的可视化从始至终最重要的目的是为了向读者展现数据的内在联系，无论是世界各地区与销售额的关系反应出的全球大环境下的游戏销售趋势亦或者是各大游戏厂商销售额的变化与游戏类型、游戏平台的发展变迁分析都是为了体现数据背后的现实意义，这是本文最有价值的地方。

# 根据可视化的结论分析与展望

5.1结论分析

现在让我们带着前文的可视化成果来分析一开始我们提到的几个问题从而得出结论：

1. 全球各地区的游戏销售变化趋势如何？为什么会这样？

尽管数据由于一些缺失导致2017年之后的数值显示不明显，但可以明确的看出电子游戏发行量在2010年之前迅猛发展，特别是在1995年至2009年，呈现指数式爆发增长。之后发行量逐渐下降，并且一年的销售量通常与给定年份发布的游戏成比例。不难猜想发行量下降一方面是电子游戏在高速发展后遭遇到手游市场冲击难免后劲不足，另一方面2008年全球经济危机以及2019年全球疫情的影响导致后续发行量快速下降乃至后期数据过少不足以分析。还有一种可能智能手机的引入和互联网的普及。由于该数据集只包含著名的基于平台的游戏数据，而不包含flash、mobile和许多其他由小型出版商开发的游戏，因此我们不能妄下论断去平均视频游戏行业在2008年开始呈下滑趋势。

图表, 直方图

描述已自动生成

图表

描述已自动生成

但有一点可以明确，2010年后游戏的发布量变少，可能是由于先前发布的游戏只需要很小的改变或是只需要通过一些更新来完成。2009/2010年是电子游戏的黄金时代。不仅是对平台背后的公司，或是对产品负责的工作室，还有对消费者。在2009年，像《使命召唤现代战争2》、《小星球》（psp）和《刺客信条2》这样的大片游戏都是非常好的游戏，在2009年的发行季之前，市场营销非常好。

从上面的图表中，我们还可以推断一个地区的哪些公民在电子游戏上花了多少钱。显而易见，北美人在电子游戏上的花费大约是欧洲人的两倍，是日本人的四倍。

1. 全球受欢迎的游戏发行商有哪些？为什么会如此受欢迎？

我认为，在平台市场上，任天堂是一个强大的平台。在这部分分析中，这似乎是一个反复出现的主题。任天堂已经年复一年地证明，他们的游戏质量是一流的。另一方面Activision展示了制作高质量游戏的能力。随着《使命召唤》每年发布，年轻一代不再那么在意故事情节和单人游戏模式，在线领域则被activision和EA主导。

图表, 瀑布图

描述已自动生成

图表

描述已自动生成

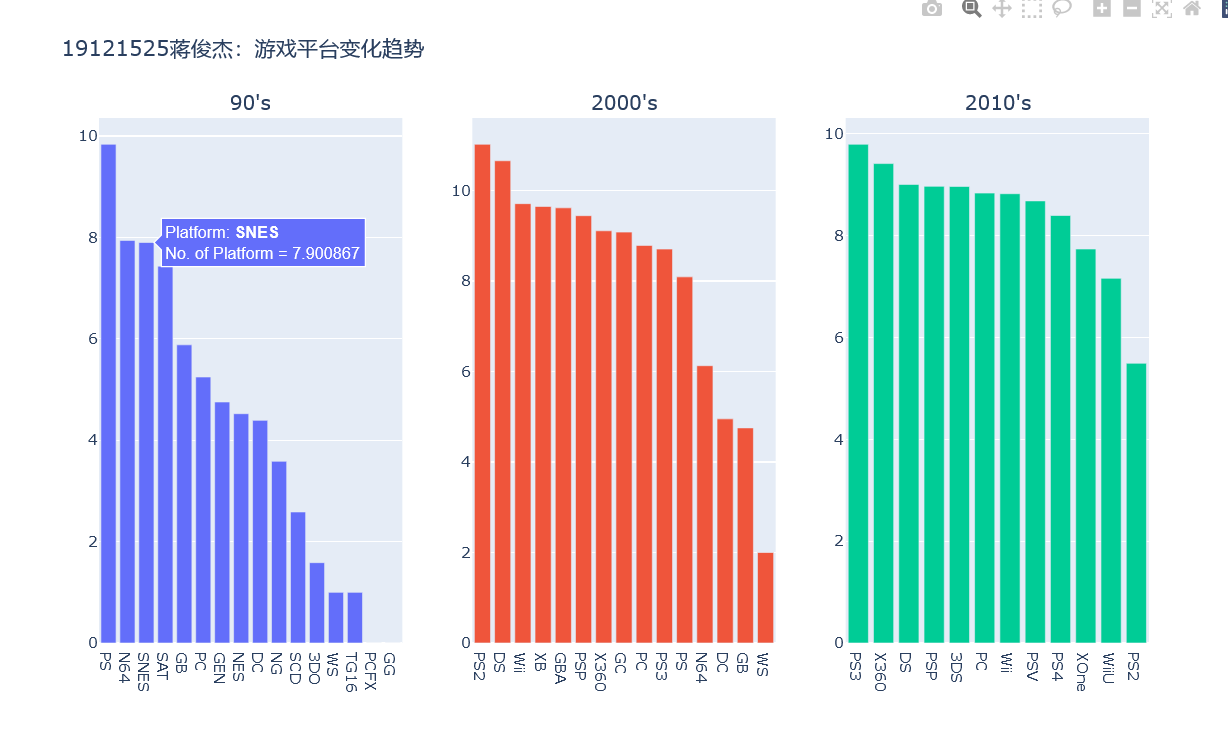
在分析游戏厂商受欢迎程度时本文绘制了游戏发售量与营销额的柱线图，可以惊讶的发现，其实在游戏的发布数量上并不是特别突出的但在游戏销售上却占比非常的高。这或许与它追求高质量游戏的目标有关，而除此之外还有别的因素吗？这就引入了第三部分的问题分析。

1. 游戏平台和游戏类型的变化趋势？

这些图表显示了过去40年中每种类型开发的游戏数量。游戏发行商根据市场需求开发游戏，在给定的时间片中，每种类型中发布的游戏数量成为一个有用的指标，可以显示类型的受欢迎程度，这对了解市场的动态和不断变化的玩家偏好有重要的意义。可以看到，除了上世纪90年代，动作片仍然是最受欢迎的类型。动作和体育从一开始就处于榜首，而益智游戏在过去三十年中一直是最不受欢迎的游戏。角色扮演游戏受欢迎程度的上升趋势显而易见，这表明玩家想要玩能让他们在虚拟世界中“成为另一个人”的游戏。这是一个重要的结论，可能有助于游戏出版商了解视频游戏行业的发展方向。观察历年来游戏平台的变化情况，主流游戏平台由PS转向PS2，进而变成PS2、DS、WII，逐步发展为PS3、DS、WII、X360。

图表, 条形图

描述已自动生成



另外我们可以再根据雷达图来了解为什么任天堂在视频电子游戏领域占据着如此重要的地位，可以通过下图看到任天堂在最受欢迎的六大游戏类型中销售额方面都占据着领先和重要地位。尤其是在近年来飞速发展的角色扮演类游戏中独占鳌头。

图表, 雷达图

描述已自动生成

5.2 未来展望

通过以上对给定视频游戏销售数据集的详尽分析，出版商在发布新游戏时必须记住以下几点：北美是最有利可图的地区，因此，如果你选择在那里开展游戏，成功的机会会更大。动作、体育和射击游戏是全世界最受欢迎的游戏类型，除了日本，角色扮演游戏最受欢迎。因此，必须根据特定地区的偏好，在特定地区推出相应的游戏。此外，人们还可以看到角色扮演视频游戏的受欢迎程度呈上升趋势，因此出版商必须专注于开发角色扮演游戏，使玩家能够参与社区并获得虚拟社交体验，从而获得流行和成功。平均而言，一款电子游戏的收入约为50万美元。因此，为了盈利，人们必须据此设计制作游戏的成本。

本文的可视化交互或许还存在着不足，但在数据之间相关信息的分析以及选择绘图方式上花费了较多的时间，在这里需要指出本文的绘图方式有echarts,python两种方式，都在代码连接中给出。在之后的学习中需要对可视化交互方面的知识做更深一步的了解。