

项目 3 旅游业数据图形展示

教学目标

知识目标

- ◆ 了解 Seaborn 库的基本概念和功能。
- ◆ 学习如何安装和配置 Seaborn 库。
- ◆ 掌握绘制计数图、箱线图、柱形图等图形的方法。
- ◆ 掌握绘制回归关系图、小提琴图、折线图等图形的方法。
- ◆ 掌握绘制直方图、饼图、散点图等图形的方法。

能力目标

- ◆ 能够运用 Seaborn 库绘制各种类型的数据图形。
- ◆ 能够根据实际需求选择合适的图形来展示数据。
- ◆ 能够对图形进行美化和调整，使其更具吸引力和可读性。

素质目标

- ◆ 具备优秀的数据分析能力。
- ◆ 具备独立学习和解决问题的能力。
- ◆ 具备团队合作和沟通能力。

教学过程

（一）项目导入

1. 介绍项目背景

通过讲述旅游业在国家经济发展中的重要地位，引出对旅游业数据进行图形展示的必要性和重要性。强调了解国内旅游业市场的游客分布、消费趋势和旅游景点偏好，以及国际旅游业收入及构成等数据的重要性。

2. 展示项目目标

明确本项目将根据我国旅游数据，使用 Seaborn 生成图形，直观地展示近年来我国国内游客分布对比、外国入境游客事由分布趋势及入境人数和国际旅游外汇收入关系。

（二）知识讲解

Seaborn 简介

Seaborn 介绍: 详细讲解 Seaborn 是建构在 Matplotlib 基础上的 Python 数据可视化库,其设计理念强调美观性,提供了更高层次的接口,使用户能够轻松地生成专业水平的图形。

Seaborn 特点: 通过具体示例,介绍 Seaborn 的美观默认样式、统计数据可视化功能、内置主题和颜色调色板、适用于数据框架、高级绘图功能、可视化统计模型结果和强大的色彩控制等特点。

Seaborn 与其他可视化库的对比: 对比 Seaborn 与 Matplotlib、Plotly、Altair 等常见可视化库的优缺点,让学生了解 Seaborn 在不同场景下的适用性。

seaborn 自带数据集: 介绍 Seaborn 内置的示例数据集,如 anscombe 数据集、tips 数据集、flights 数据集、iris 数据集等,并演示如何使用 `seaborn.load_dataset()` 函数加载数据集。

Seaborn 安装与配置

安装 Seaborn: 讲解在 cmd 或者终端中使用 pip 安装 Seaborn 的方法,包括普通安装、安装到用户目录和安装特定版本的命令。

配置 Seaborn: 演示如何导入 Seaborn 库,设置图形风格(如 whitegrid、darkgrid 等),调整图形尺寸,以及设置颜色调色板和字体大小等配置选项。

示例代码讲解: 通过一个简单的 Seaborn 示例代码,详细讲解代码的每个步骤,包括导入库、设置图形风格和尺寸、加载数据集、绘制散点图、添加标题和标签等,让学生熟悉 Seaborn 的基本使用流程。

（三）任务实践

任务 3 - 1: 国内游客人次数据展示

任务描述: 介绍本任务要求选用 seaborn 模块中合适的图形函数,展示城镇居民国内游客人次和农村居民国内游客人次的变化趋势。

绘图方法讲解: 分别讲解计数图、箱线图和柱形图的概念、适用场景和绘制方法,包括函数的完整形式和常用参数。

在 Seaborn 中,计数图(countplot)是一种用于显示分类变量的频率分布的图形。它简单地计算数据集中每个类别的出现次数,并将结果可视化为条形图或柱形图。计数图通常用

于探索数据集中各个类别的分布情况，以便快速了解各类别的相对频率。

Seaborn 中用于绘制计数图的主要函数是 `countplot()` 函数，以下是 `countplot` 函数的完整形式及一些常用参数：

```
sns.countplot(x=None, hue=None, data=None, order=None, hue_order=None, orient=None, color=None, palette=None, saturation=0.75, dodge=True, ax=None, **kwargs)
```

箱线图 (Boxplot)，也被称为盒须图、盒式图，是一种用于显示数据分布情况的统计图。它主要展示了一组数据的中位数、上下四分位数、最大值、最小值以及异常值（离群点）。

箱线图的绘制过程包括以下几个要素：

箱体 (Box)： 画出上下四分位数 (Q1 和 Q3) 之间的矩形，中间的线表示中位数 (Q2)。

须 (Whisker)： 从箱体延伸出来的线段，表示数据的范围。通常，须的长度由数据的分布情况确定，可以是数据的最小值和最大值，也可以根据一定的标准计算。

异常值 (Outliers)： 在图表上以点的形式标出的数据点，它们偏离了常规的数据分布。异常值的定义通常是根据一定的统计学标准，例如，距离四分位距离超过 1.5 倍的标准差。

箱线图能够直观地显示数据的中心趋势和离散程度，有助于识别数据的异常值和整体分布情况。

Seaborn 中用于绘制箱线图的主要函数是 `boxplot()` 函数，以下是 `boxplot` 函数的完整形式及一些常用参数：

```
seaborn.boxplot(x=None, y=None, hue=None, data=None, order=None, hue_order=None, orient=None, color=None, palette=None, saturation=0.75, width=0.8, dodge=True, fliersize=5, linewidth=None, whis=1.5, notch=False, ax=None, **kwargs)
```

柱形图，又称长条图、柱状统计图、条图、条状图、棒形图，是一种以长方形的长度为变量的[统计图表](#)。长条图用来比较两个或以上的价值（不同时间或者不同条件），只有一个变量，通常利用于较小的[数据集](#)分析。长条图亦可横向排列，或用多维方式表达。

Seaborn 中用于绘制散点图的主要函数是 `barplot()`，以下是 `barplot` 函数的完整形式及一些常用参数：

```
seaborn.barplot(x=None, y=None, hue=None, data=None, estimator=np.mean, ci=95, orient=None, color=None, palette=None, dodge=True)
```

1. **示例代码分析：**通过具体的示例代码，详细分析如何使用 Seaborn 绘制计数图、箱线图

图和柱形图，包括数据准备、函数调用、图形属性设置等步骤。

2. **实践操作：**学生根据示例代码，读取 “data.xls” 表中 “国内旅游情况” 表单数据，并根

据历年 “城镇居民国内游客 (百万人次)” 和 “农村居民国内游客 (百万人次)” 行数据绘制柱形图。教师巡视指导，及时解决学生遇到的问题。

任务 3 - 2：外国入境游客人次按年龄展示（1.5 课时）

1. **任务描述：**介绍本任务要求选用 seaborn 模块中合适的图形函数，展示外国入境游客

（按照年龄）变化趋势。

2. **绘图方法讲解：**分别讲解回归关系图、小提琴图和折线图的概念、适用场景和绘制方法，

包括函数的完整形式和常用参数。

Seaborn 中的回归关系图是一种用于可视化两个变量之间关系的图表，通常用于显示两个连续变量之间的线性关系，并估计它们之间的线性回归模型。这种图表可以帮助我们观察两个变量之间的趋势、相关性以及可能存在的线性关系。

Seaborn 中用于绘制回归关系图的主要函数是 `regplot()` 函数，以下是 `regplot()` 函数的完整形式及一些常用参数：

```
sns.regplot(x, y, data=None, scatter_kws=None, line_kws=None)
```

在 Seaborn 中，小提琴图 (Violin Plot) 是一种用于可视化数值数据分布的统计图表。它结合了箱线图和核密度估计图的特点，能够显示数据的中位数、四分位数、离群值以及整体数据分布的形状。

小提琴图的主要组成部分包括：

箱线图：显示数据的中位数、四分位数和离群值。

核密度估计曲线：通过曲线的宽度和高度反映了数据在不同数值范围内的密度分布情况。

数据点：在小提琴图的两侧显示了每个观测值的分布情况。

小提琴图通常用于比较不同类别或组之间数值数据的分布情况，以便直观地了解数据的

形状和分布特征。Seaborn 提供了 `violinplot()` 函数来绘制小提琴图，并且支持对数据进行分组、添加色调等功能，使得小提琴图的应用更加灵活和多样化。

Seaborn 中用于小提琴图的主要函数是 `violinplot()`，以下是 `violinplot` 函数的完整形式及一些常用参数：

```
sns.violinplot(x=None, y=None, hue=None, data=None, order=None, hue_order=None,
bw='scott', cut=2, scale='area', scale_hue=True, gridsize=100, width=0.8, inner='box',
split=False, dodge=True, orient='v', color=None, palette=None, saturation=0.75, ax=None,
**kwargs)
```

折线图是一种用于显示数据随着连续变量的变化而变化的图形。在折线图中，数据点通过连接线条来展示趋势、变化或关系。折线图常用于表示时间序列数据，显示变量随时间的演变。它能够清晰地展示数据的趋势和周期性。

Seaborn 中用于折线图的主要函数是 `lineplot()`，以下是 `lineplot()` 函数的完整形式及一些常用参数：

```
seaborn.lineplot(x=None, y=None, hue=None, style=None, size=None, data=None,
palette=None, markers=True, dashes=True, legend="auto",
hue_order=None)
```

3. **示例代码分析：**通过具体的示例代码，详细分析如何使用 Seaborn 绘制回归关系图、

小提琴图和折线图，包括数据准备、函数调用、图形属性设置等步骤。

4. **实践操作：**学生根据示例代码，读取“data.xls”表中“按性别、年龄和事由分外国入境游客”表单数据，并根据历年“14 岁以下外国人入境游客（万人次）”、“15 - 24 岁外国人入境游客（万人次）”、“25 - 44 岁外国人入境游客（万人次）”、“45 - 64 岁外国人入境游客（万人次）”、“65 岁以上外国人入境游客（万人次）”行数据绘制折线图。教师巡视指导，及时解决学生遇到的问题。

任务 3 - 3：旅游业发展数据展示

1. **任务描述：**介绍本任务要求选用 seaborn 模块中合适的图形函数，展示外国人入境游

客和国际旅游外汇收入之间的关系。

2. **绘图方法讲解：**分别讲解直方图和散点图的概念、适用场景和绘制方法，包括函数的完

整形式和常用参数。

直方图是一种统计图表，用于展示数据的分布情况。它将数据分成若干个等宽的区间（通常称为“箱子”或“柱子”），然后统计每个区间内数据的频数或频率。直方图的横轴通常表示数据的取值范围，纵轴则表示每个区间内数据的频数或频率。通过绘制这些区间的条形图，直方图能够直观地展示数据的分布特征，包括数据的中心趋势、离散程度和异常值等。直方图通常用于探索性数据分析，帮助人们理解数据的整体分布情况，并发现数据中的模式和规律。

Seaborn 中用于小提琴图的主要函数是 `histplot()`，以下是 `histplot` 函数的完整形式及一些常用参数：

```
seaborn.histplot(x=None, y=None, hue=None, multiple='layer', binwidth=None,
binrange=None, bins=10, kde=False, stat='count', common_bins=True, element='bars',
fill=None)
```

散点图是一种用于显示两个数值变量之间关系的图形。它通过在二维平面上以点的形式表示数据点，其中每个点的 X 轴和 Y 轴坐标分别对应两个变量的值。散点图可用于观察变量之间的趋势、集中程度、异常值以及变量之间的相关性。

Seaborn 中可以使用 `scatterplot()` 函数绘制散点图，以下是 `scatterplot` 函数的完整形式及一些常用参数：

```
seaborn.scatterplot(x=None, y=None, hue=None, style=None, size=None, data=None,
palette=None, hue_order=None, hue_norm=None, sizes=None)
```

3. **示例代码分析：**通过具体的示例代码，详细分析如何使用 Seaborn 绘制直方图和散点

图，包括数据准备、函数调用、图形属性设置等步骤。

4. **实践操作：**学生根据示例代码，读取 “data.xls” 表中 “旅游业发展情况” 表单数

据，并根据历年 “入境游客人数（万人次）”、“国际旅游外汇收入（百万美元）” 数据绘制散点图。教师巡视指导，及时解决学生遇到的问题。

