

Python 数据可视化 课后习题

项目 1 数据可视化导论

一、选择题

1. 下列选项中，关于数据可视化描述错误的是()。
 - A. 数据可视化可以简单地理解为将不易描述的事物形成可感知画面的过程
 - B. 数据可视化的目的是准确、高效、全面地传递信息
 - C. 数据表格是数据可视化最基础的应用
 - D. 数据可视化对后期数据挖掘具有深远的影响
2. 关于常见的图表，下列描述正确的是()。
 - A. 柱形图可以反映数据增减的趋势
 - B. 条形图是横置的直方图
 - C. 饼图用于显示数据中各项大小与各项总和的比例
 - D. 雷达图是一种可以展示多变量关系的图表
3. 下列图表中，可以反映 3 个变量之间关系的是()。
 - A. 折线图
 - B. 柱形图
 - C. 散点图
 - D. 气泡图
4. 下列哪个可视化库可以生成 ECharts 图表?()
 - A. matplotlib
 - B. seaborn
 - C. Bokeh
 - D. pyecharts
5. 下列选项中，属于数据之间逻辑关系的是()。(多选)
 - A. 比较
 - B. 分布
 - C. 构成
 - D. 联系

二、填空题

1. 数据可视化是从数据空间到_____空间的映射。
2. 狭义的数据可视化是将数据以_____、图形、地图的方式进行呈现。
3. _____是一种利用数据中的 5 个统计量描述数据的图表。
4. matplotlib 是一个主要用于绘制_____图表的 Python 库。
5. _____提供了包管理器、环境管理器，包括诸如 numpy、pandas、matplotlib、scipy 等 180 多个科学计算包及其依赖项。

三、判断题

1. 数据可视化是一个抽象的过程。()
2. 散点图可以清晰地展示数据增减的趋势、速率、规律、峰值等特征。()
3. 柱形图与直方图展示的效果完全相同。()

4.matplotlib 只能采用面向对象的方式开发程序。()

四、简答题

- 1.什么是数据可视化?数据可视化的作用是什么?
- 2.列举一些常用的数据可视化工具和软件。
- 3.常见的数据可视化图表有哪些?简述各种图表的特点和适用场景。

项目 2 人口普查数据图形展示

一、选择题

1、下列代码中，绘制散点图的是 ()。

- A: plt.scatter(x,y)
- B: plt.plot(x,y)
- C: plt.legend('upper left')
- D: plt.xlabel('散点图')

2、下列函数中，用于保存当前生成的图表的是 ()。

- A: figure()
- B: hist()
- C: savefig()
- D: show()

3、下列选项中，可以一次性创建多个子图的是 ()。

- A: figure()
- B: subplot()
- C: add_subplot()
- D: subplots()

4、在创建 Figure 对象时，可以指定哪个参数来设置画布的尺寸?

- A: num
- B: dpi
- C: figsize
- D: facecolor

二、填空题

- 1、_____使用一系列高度不等的纵向条纹或线段表示数据分布的情况。
- 2、饼图可以显示一个数据序列中各项的大小与各项总和的_____。
- 3、Figure 对象允许划分为多个绘图区域，每个绘图区域都是一个_____ 对象，被称为子图。
- 4、如果不希望在默认的画布上绘制图形，则可以调用 _____ 函数构建一张新的空白画布。

三、判断题

- 1、Figure 对象可以理解为一块空白的画布，用于容纳图表的各种组件。
- 2、Figure 对象允许划分为多个绘图区域。
- 3、subplot(3,2,3)和 subplot(3,2,3)是等价的。
- 4、位于左上角的子图编号为 1。
- 5、每调用一次 add_subplot()方法会规划画布划分子图，且只会添加一个子图。

四、编程题

- 1、读取“人口数据.xlsx”中“按受教育程度分人口数”表单数据，提取各受教育程度人数的列数据，绘制饼图。
- 2、读取“人口数据.xlsx”中“人口年龄结构抚养比”表单数据，并根据历年各年龄段列数据绘制茎叶图。
- 3、读取“人口数据.xlsx”中“总人口”表单数据，根据历年“城镇人口”和“乡村人口”数据绘制堆叠区域图。
- 4、在一个画布上同时呈现以上三个图表。

项目 3 旅游业的数据图形展示

一、选择题

Seaborn 是构建在哪个库基础上的 Python 数据可视化库？（ ）

- A. NumPy
- B. Pandas
- C. Matplotlib
- D. Plotly

以下哪个函数用于在 Seaborn 中绘制计数图？（ ）

- A. scatterplot()
- B. countplot()
- C. boxplot()
- D. barplot()

要绘制一个展示两个连续变量之间线性关系的回归关系图，应使用 Seaborn 中的哪个函数？（ ）

- A. regplot()
- B. violinplot()
- C. lineplot()
- D. histplot()

Seaborn 中 pairplot() 函数默认会为数据集中的每一对变量创建什么？（ ）

- A. 箱线图矩阵
- B. 散点图矩阵
- C. 折线图矩阵
- D. 直方图矩阵

若要绘制一个显示数据分布情况，结合了箱线图和核密度估计图特点的图表，应使用 Seaborn 中的哪个函数？（ ）

- A. boxplot()
- B. violinplot()
- C. swarmplot()
- D. stripplot()

二、填空题

Seaborn 提供了更高层次的接口，其默认样式比 Matplotlib 更_____。

在 Seaborn 中，barplot() 函数的 estimator 参数默认值是_____。

若要加载 Seaborn 自带的“tips”数据集，可使用函数_____。

绘制折线图时，Seaborn 中使用的主要函数是_____。

为了避免分类散点图中数据点的重叠，可使用 Seaborn 中的_____函数。

三、代码填空题

以下代码用于绘制一个显示不同种类鸢尾花的花萼长度分布的箱线图，请补充完整。

```
import seaborn as snsimport matplotlib.pyplot as plt
# 加载鸢尾花数据集
iris = sns.load_dataset('iris')
# 设置图形大小
plt.figure(figsize=(10, 6))
# 绘制箱线图
sns.boxplot(x=_____, y=_____, data=iris)
# 添加标题和标签
plt.title('不同种类鸢尾花的花萼长度分布')
plt.xlabel('鸢尾花种类')
plt.ylabel('花萼长度 (cm)')
# 显示图形
plt.show()
```

四、简答题

简述 Seaborn 与 Matplotlib 的主要区别。

解释箱线图 (Boxplot) 的主要组成部分及其作用。

说明 Seaborn 中 histplot() 函数的作用及主要参数的含义。

五、数据分析与绘图题

1. 读取“data.xls”表中“国内旅游情况”表单数据，使用 Seaborn 绘制柱形图，展示历年“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)”的数据对比。
2. 读取“data.xls”表中“旅游业发展情况”表单数据，使用 Seaborn 绘制散点图，展示“入境游客人数 (万人次)”和“国际旅游外汇收入 (百万美元)”之间的关系。

项目 4 人们生活的数据图形展示

一、选择题

- 1、以下哪种图表最适合展示不同地区居民收入的中位数和离散程度 ()
- A. 折线图
 - B. 箱形图

- C. 饼图
- D. 漏斗图

2、在 pyecharts 中，用于组合多个图表在同一页面并列显示的容器类是（ ）

- A. Grid
- B. Tab
- C. Timeline
- D. Page

3、在 pyecharts 中，控制图表整体主题样式的配置项是：

- A. TitleOpts
- B. InitOpts
- C. TooltipOpts
- D. LegendOpts

4、下列哪种图表能直观展示消费流程中各环节的转化率（ ）

- A. 雷达图
- B. 漏斗图
- C. 散点图
- D. 矩形树图

5、矩形树图（Treemap）的核心作用是：

- A. 比较多个类别的数值大小
- B. 展示层次结构数据的占比关系
- C. 分析两个变量的相关性
- D. 显示时间序列的动态变化

二、填空题

- 1、在 pyecharts 中，初始化图表尺寸的配置项参数是 InitOpts(____=800, height=500)。
- 2、绘制饼图时，通过 rosetype="_____" 参数可生成南丁格尔玫瑰图。
- 3、在箱形图中，箱体上下边界分别表示数据的第_____和第_____四分位数。
- 4、使用 Tab 组件时，每个子图表通过 _____ 方法添加到选项卡。
- 5、全局配置项 ToolboxOpts 默认包含的功能按钮包括：数据视图、____、还原、保存图片。

三、判断题

- 1、矩形树图（Treemap）的每个区块面积必须与数据值成比例。（ ）
- 2、set_global_opts 中设置的标题会覆盖 InitOpts 中的主题配置。（ ）
- 3、set_series_opts 用于设置单个数据系列的样式。（ ）
- 4、散点图适合展示两个连续变量之间的非线性关系。（ ）
- 5、时间线轮播图仅支持按年份顺序播放。（ ）

四、编程题

- 1、读取“城镇居民人均收入情况.xls”中“城镇居民人均收入”数据，绘制柱形图展示 2014-2023

年 10 年间城镇居民人均收入情况。

2、读取“城镇居民人均收入情况.xls”中“城镇居民人均收入”数据、“农村居民人均收入情况.xls”中“农村居民人均收入”数据、“全国居民人均收入情况.xls”中“全国居民人均收入”数据，绘制折线图展示 2014-2023 年 10 年人均收入情况。

3、读取“城镇居民人均支出情况.xls”和“农村居民人均支出情况.xls”中食品烟酒、衣着、居住、交通通信、教育文化娱乐和医疗保健方面数据，绘制矩形树图，展示 2023 年城镇居民和农村居民人均支出情况。

4、选中上述中任意两个图表绘制层叠多图或者并行多图。

Python 数据可视化 课后习题参考答案

项目 1 数据可视化导论

一、选择题

- 1.B
- 2.C
- 3.D
- 4.D
- 5.BCD

三、 填空题

- 1. 图形
- 2. 图表
- 3. 箱线图
- 4. 2D 或 3D
- 5. anaconda

三、判断题

- 1. ✓
- 2. ×
- 3. ×
- 4. ×

四、简答题

1.

数据可视化，顾名思义，就是将数据转换成图或表等，以一种更直观的方式展现和呈现数据，让读者能“一眼看懂”你想表达的信息。通过“可视化”的方式，复杂的数据通过图形化的手段进行有效表达，准确高效、简洁全面地传递某种信息，甚至我们帮助发现某种规律和特征，挖掘数据背后的价值。

2.

Matplotlib Seaborn pyecharts

3.

折线图以折线的方式形象地反映事物沿某个维度的变化趋势，能够清晰地展示数据增减的趋势、速率、规律及峰值等特征。

柱形图是由一系列宽度相等的纵向矩形条组成的图表，它利用矩形条的高度表示数值以此反映不同分类数据之间的差异。

堆积面积图是由若干折线与折线或水平坐标轴之间的填充区域组成的图表

直方图又称质量分布图，是由一系列高低不等的纵向矩形，条或线段组成的图表，用于反映数据的分布和波动情况。

箱形图又称盒须图、箱线图，是一种利用数据中的 5 个统计量(最小值、下四分位数中位数、上四分位数和最大值)描述数据的图表，主要用于反映一组或多组数据的对称性分布程度等信息

项目 2 人口普查数据图形展示

一、选择题

- 1、A
- 2、C
- 3、D
- 4、C

二、填空题

- 1、直方图
- 2、百分比
- 4、axes
- 5、figure

三、判断题

- 1、√
- 2、√
- 3、√
- 4、√
- 5、√

四、编程题

略

项目 3 旅游业的数据图形展示

一、选择题答案

- C
B
A
B
B

二、填空题答案

美观

`np.mean`

`sns.load_dataset("tips")`

`lineplot()`

`swarmplot()`

三、代码填空题答案

`import seaborn as sns`

`import matplotlib.pyplot as plt`

`# 加载鸢尾花数据集`

`iris = sns.load_dataset('iris')`

```
# 设置图形大小
plt.figure(figsize=(10, 6))
# 绘制箱线图
sns.boxplot(x='species', y='sepal_length', data=iris)
# 添加标题和标签
plt.title('不同种类鸢尾花的花萼长度分布')
plt.xlabel('鸢尾花种类')
plt.ylabel('花萼长度 (cm)')
# 显示图形
plt.show()
```

四、简答题答案

简述 Seaborn 与 Matplotlib 的主要区别：

高级接口：Seaborn 提供更高层次接口，能轻松创建统计图表；Matplotlib 更注重绘图灵活性和底层控制。

美观性：Seaborn 默认样式更美观，无需额外设置可得专业外观图形；Matplotlib 默认样式相对简单。

颜色调色板：Seaborn 提供丰富颜色调色板突出图形不同元素；Matplotlib 需手动设置颜色。

解释箱线图（Boxplot）的主要组成部分及其作用：

箱体（Box）：由上下四分位数（Q1 和 Q3）间的矩形表示，中间线为中位数（Q2），展示数据中间 50% 分布范围。

须（Whisker）：从箱体延伸的线段，代表数据范围，其长度依数据分布确定，可展示数据整体分布跨度。

异常值（Outliers）：以点形式标出，偏离常规数据分布，通常按距离四分位距离超 1.5 倍标准差定义，用于识别数据极端值。

说明 Seaborn 中 histplot() 函数的作用及主要参数的含义：

作用：绘制直方图，展示数据分布情况。

主要参数含义：

x：指定 X 轴数据。

y：指定 Y 轴数据。

hue：对数据分组，依不同取值绘制不同颜色直方图。

multiple：处理多个观测值数据，默认 'layer'，即层叠直方。

binwidth：指定直方图条形宽度。

binrange：指定直方图条形范围。

bins：指定直方图条形数量。

kde：是否显示核密度估计曲线。

stat：确定绘图统计量，默认 'count'，即绘制计数直方图。

10.common_bins：是否所有组用相同箱子，默认 True。

11.element：确定绘图元素类型，默认 'bars'，即绘制条形直方图。

12.fill：确定是否填充直方图条形颜色，默认 None，即不填充。

五、数据分析与绘图题答案

读取 “data.xls” 表中 “国内旅游情况” 表单数据，使用 Seaborn 绘制柱形图，展示历年 “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)” 的数据对比：

```
import pandas as pd
import seaborn as sns
```

```

import matplotlib.pyplot as plt

# 设置 matplotlib 支持中文显示
plt.rcParams['font.sans-serif'] = ['SimHei'] # 使用黑体字体，可根据系统情况替换为其他中文字体，如 'Microsoft YaHei'
# 解决负号显示问题
plt.rcParams['axes.unicode_minus'] = False

# 读取数据
data = pd.read_excel('data.xls', sheet_name='国内旅游情况')

# 提取数据
subset = data.iloc[1:3, 1:]
# 转换数据格式
melted_data = subset.T.reset_index()
melted_data.columns = ['年份', '城镇居民国内游客', '农村居民国内游客']
melted_data = pd.melt(melted_data, id_vars='年份', var_name='游客类型', value_name='国内游客（百万人次）')
# 设置图形大小
plt.figure(figsize=(12, 6))
# 绘制柱形图
sns.barplot(x='年份', y='国内游客（百万人次）', hue='游客类型', data=melted_data)
# 添加标题和标签
plt.title('城镇居民和农村居民国内游客数据对比')
plt.xlabel('年份')
plt.ylabel('国内游客（百万人次）')
# 调整 X 轴标签角度
plt.xticks(rotation=45)
# 显示图形
plt.show()

```

2. 读取 “data.xls” 表中 “旅游业发展情况” 表单数据，使用 Seaborn 绘制散点图，展示 “入境游客人数（万人次）” 和 “国际旅游外汇收入（百万美元）” 之间的关系：

```

import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# 设置显示中文字体
plt.rcParams['font.sans-serif'] = ['SimHei']
plt.rcParams['axes.unicode_minus'] = False

# 读取数据
data = pd.read_excel('data.xls', sheet_name='旅游业发展情况')
# 提取数据
tourist_arrivals = data.iloc[2, 1:]
foreign_exchange_income = data.iloc[10, 1:]

# 创建 DataFrame

```

```
df = pd.DataFrame({
    '入境游客人数（万人次）': tourist_arrivals,
    '国际旅游外汇收入（百万美元）': foreign_exchange_income})
# 设置图形大小
plt.figure(figsize=(10, 6))
# 绘制散点图
sns.scatterplot(x='入境游客人数（万人次）', y='国际旅游外汇收入（百万美元）', data=df)
# 添加标题和标签
plt.title('入境游客人数与国际旅游外汇收入关系')
plt.xlabel('入境游客人数（万人次）')
plt.ylabel('国际旅游外汇收入（百万美元）')
plt.show()
```

项目 4 人民生活数据图形展示

一、选择题

- 1、B
- 2、A
- 3、B
- 4、B
- 5、B

二、填空题

- 1、width
- 2、area
- 3、25%、75%
- 4、add()
- 5、缩放

三、判断题

- 1、√
- 2、×
- 3、√
- 4、√
- 5、×

四、编程题

略