**Python 儿童身体健康情况 可视化分析**

**1. 明确项目需求与目标**

**任务描述：**

联合国儿童基金会曾发表文章指明发展中国家 5 岁以下儿童中超过四分之一体重严重 不足，而全球每年有 560 万儿童因营养不良死亡，而儿童营养不良的现象受到大规模贫困、 医疗系统薄弱、社会动荡、武装冲突和气候变化等多重因素影响。如 2020 年 9 月发布新闻， 西非萨赫勒有数百万儿童受到营养不良问题的严重影响，该地区的儿童在出生后面临喂养 不足、卫生状况恶劣、优质医疗服务有限和安全用水难以获取等生存挑战，而新冠疫情及其 所引发的社会经济危机则使得情况进一步恶化。营养不良和感染之间的相互作用可能造成疾病恶化和营养状况恶化的潜在致命循环。 联合国儿童基金会执行主任安东尼·莱克说，这些严重营养不良和正在发生的饥荒大部分是 人为原因导致的，国际社会需要尽快行动起来，拯救更多生命。

**任务分析：**

本项目旨在利用可视化分析方法，直观地展示 2000 年至 2020 年儿童身体情况的发展变化，跟踪近 20 年来儿童营养不良的水平和趋势。挖掘存在严重营养不良的国家，实现儿童营养不良研究报告，对儿童营养不良的估计将有助于确定世界是否正在实现可持续发展 目标——特别是具体目标 2.2，即“到 2030 年消除一切形式的营养不良”，该目标属于“消 除饥饿、实现粮食安全和改善营养，促进可持续农业”。

**2.** **数据探索和数据预处理**

**任务描述：**

探索数据数值分布的情况，对比最大值与最小值之间 的区别；

清洗数据中的缺失值、重复值、异常值等情况，

确定数值估计的方法，提取该方法的儿童营养不良情况记录数据。

**任务分析：**

对儿童营养不良情况数据进行预处理可以分为以下三个步骤。

（1）选择本次数据分析所将采用的值；

（2）对数据垃圾进行清洗；

（3）选择数值估计方法；

**2-1. 选择本次数据分析所将采用的值**

通过观察数据发现原始数据中每个国家的数据都含有三种类型，分别是最大值，最小值，以及均值。其中，最大值与最小值在一年中都具有不稳定性，所以本次数据分析选择对原始数据的均值进行数据分析。具体代码如下图所示。

|  |
| --- |
| df\_sheet1\_Point\_Estimate = df\_sheet1[df\_sheet1['Estimate'].isin(['Point Estimate'])] #其他三张sheet同理 |

**2-2.对数据垃圾进行清洗**

浏览完原始数据后可以发现，原始数据存在额外多出‘Note’列，一些国家不存在数据的问题。额外多出的‘Note’列虽然某几行存在数值，但其归根到底是无用数据，而一些国家不存在数据，导致数据存在缺失值问题，其原因可能为该国家并没有被统计到，或者其数量和占比所能提供的数值过小导致，对于这两个问题都对数据的影响不大，因此直接删除即可。具体代码如下图所示。

|  |
| --- |
| df\_sheet1 = pd.read\_excel('Country\_Estimates\_2021.xlsx',sheetname=1)  df\_sheet1 = df\_sheet1.drop('Note',axis=1)  df\_sheet1.replace("-",np.nan,inplace=True)  df\_sheet1 = df\_sheet1.dropna() #其他三张sheet同理 |

**2-3.选择数值估计方法**

本次数据分析采用最常见的求均值的方法，对数值进行估计。并借此方法对儿童营养不良情况数据进行记录（包括2000 年至 2020 年各国儿童发育不良的比例，2000 年至 2020 年各国儿童发育不良的人数，2000 年至 2020 年各国儿童超重的比例，2000 年至 2020 年各国儿童超重的人数）。具体代码如下图所示。

|  |
| --- |
| Point\_Estimate = df\_sheet1\_Point\_Estimate.iloc[:,5:]  Point\_Estimate=Point\_Estimate.astype(float)  Point\_Estimate\_mean =Point\_Estimate.mean(axis=1) #其他三张sheet同理 |

**3.总体情况可视化**

**任务描述：**

统计全世界2000年至2020年儿童发育不良和超重的平均比例和平均人数，绘制图表，跟踪近二十年儿童营养不良的变动趋势以及对比 2000 年与 2020 年儿童营养不良的变化水平，根据可视化结果进行分析。

**任务分析：**

对总体情况可视化可以分为以下两个步骤。

（1）读取儿童营养不良情况数据；

（2）绘制可视化图并进行分析；

**3-1.读取儿童营养不良情况数据**

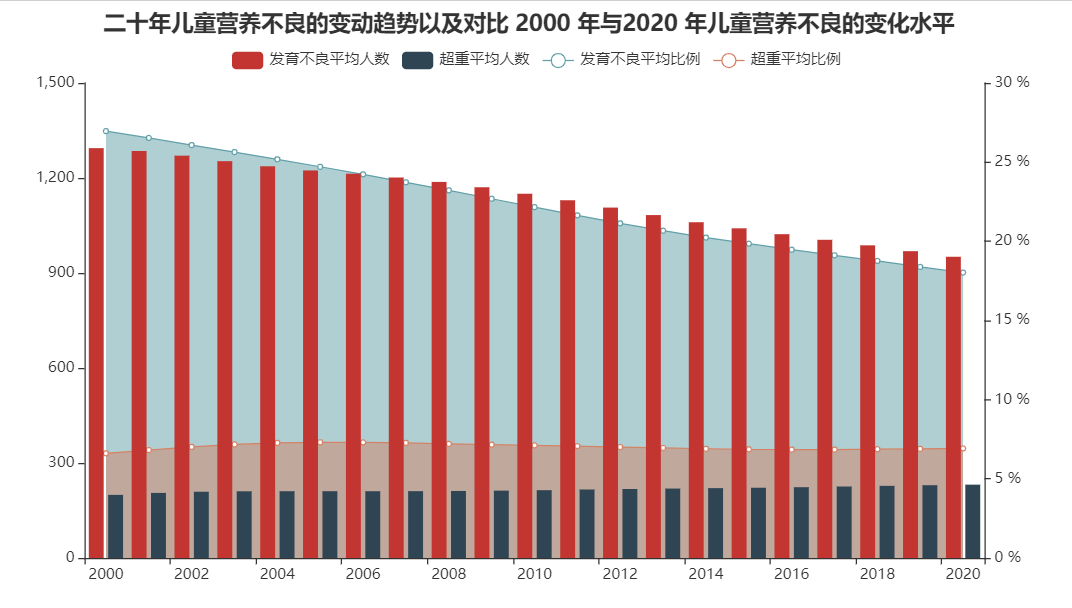
使用在预处理后的儿童营养不良平均情况的数据，将具体数值转为浮点型数据，再对每个国家从2000年到2020年每年的数据求均值。具体代码如下图所示

|  |
| --- |
| #发育不良平均比例  df\_sheet1\_Point\_Estimate = df\_sheet1[df\_sheet1['Estimate'].isin(['Point Estimate'])]  Point\_Estimate = df\_sheet1\_Point\_Estimate.iloc[:,5:]  Point\_Estimate=Point\_Estimate.astype(float)  Point\_Estimate\_mean = Point\_Estimate.agg('mean',axis=0)  #发育不良平均人数  Point\_Estimate2=Point\_Estimate2.astype(float)  Point\_Estimate\_mean2 = Point\_Estimate2.agg('mean',axis=0)  #超重平均比例  Point\_Estimate2=Point\_Estimate2.astype(float)  Point\_Estimate\_mean3 = Point\_Estimate3.agg('mean',axis=0)  #超重平均人数  Point\_Estimate2=Point\_Estimate2.astype(float)  Point\_Estimate\_mean4 = Point\_Estimate4.agg('mean',axis=0)  Point\_Estimate\_mean = Point\_Estimate\_mean.round(2)  Point\_Estimate\_mean2 = Point\_Estimate\_mean2.round(2)  Point\_Estimate\_mean3 = Point\_Estimate\_mean3.round(2)  Point\_Estimate\_mean4 = Point\_Estimate\_mean4.round(2) |

**3-2.绘制可视化图并进行分析**

本次绘图采用pyecharts的层叠图的形式展现全世界2000年至2020年儿童发育不良和超重的平均比例和平均人数。如下图所示。

|  |
| --- |
| from pyecharts import options as opts  from pyecharts.charts import Bar, Line  from pyecharts.faker import Faker  year = ['2000','2001','2002','2003','2004','2005','2006','2007','2008','2009','2010',  '2011','2012','2013','2014','2015','2016','2017','2018','2019','2020']  bar1 = (  Bar()  .add\_xaxis(year)  .add\_yaxis("发育不良平均人数", Point\_Estimate\_mean2.values.tolist())  .add\_yaxis("超重平均人数", Point\_Estimate\_mean4.values.tolist())  .extend\_axis(yaxis=opts.AxisOpts(type\_="value", position="right",min\_=0,max\_=30,axislabel\_opts=opts.LabelOpts(formatter="{value} %") ))  .set\_series\_opts(label\_opts=opts.LabelOpts(is\_show=False))  .set\_global\_opts(  title\_opts=opts.TitleOpts(title="二十年儿童营养不良的变动趋势以及对比 2000 年与2020 年儿童营养不良的变化水平",pos\_left=100),  legend\_opts=opts.LegendOpts(pos\_top = 30, pos\_right = 200)  )  )  line = (  Line()  .add\_xaxis(year)  .add\_yaxis("发育不良平均比例", Point\_Estimate\_mean.values.tolist(),yaxis\_index=1,areastyle\_opts=opts.AreaStyleOpts(opacity=0.5))  .add\_yaxis('超重平均比例',Point\_Estimate\_mean3.values.tolist(),yaxis\_index=1,areastyle\_opts=opts.AreaStyleOpts(opacity=0.5))  .set\_series\_opts(label\_opts=opts.LabelOpts(is\_show=False))  )  bar1.overlap(line)  bar1.render\_notebook() |



如图所示，可以观察到近二十年来，世界发育不良平均人数及占比逐年下降，反观近20年来世界的超重平均人数及占比则变化不大，但仍有增长的趋势。

**4.** **儿童发育不良情况可视化分析**

**任务描述：**

对各国儿童发育不良的比例和人数进行划分， 根据数值的大小区分营养不良的程度，绘制图表，统计并对比 2000 年与 2020 年不同程度分 布情况；绘制地理图，分析不同国家儿童营养不良的情况；根据不同程度，抽取排名前 3 的 国家数据，整合数据分析儿童营养不良不同程度数据的差距情况；分析对比 2000 年与 2020 年各国儿童发育不良的比例和人数变动情况。

**任务分析：**

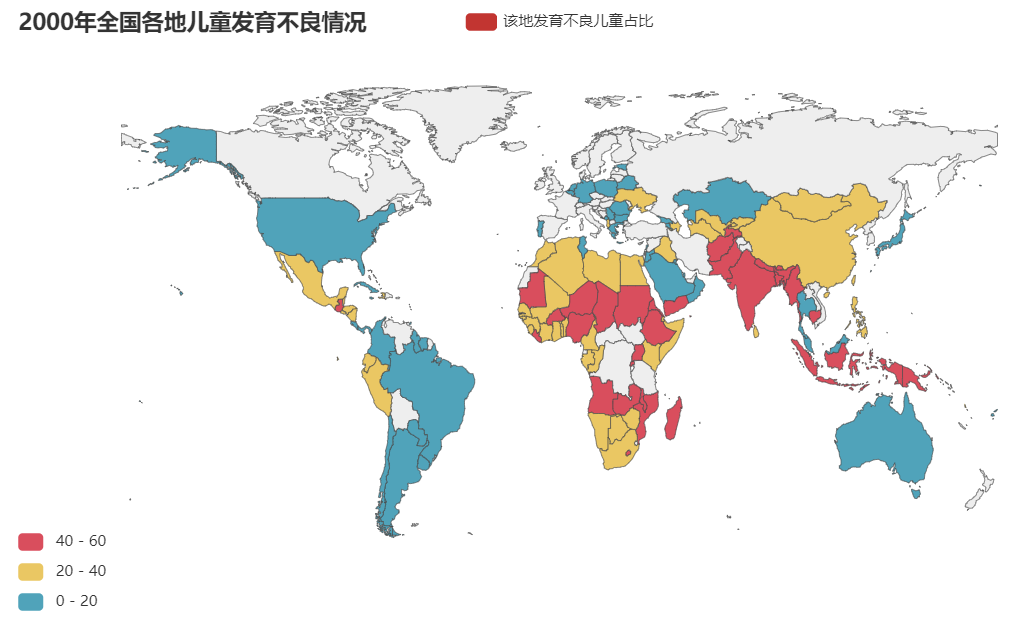
对总体情况可视化可以分为以下步骤。

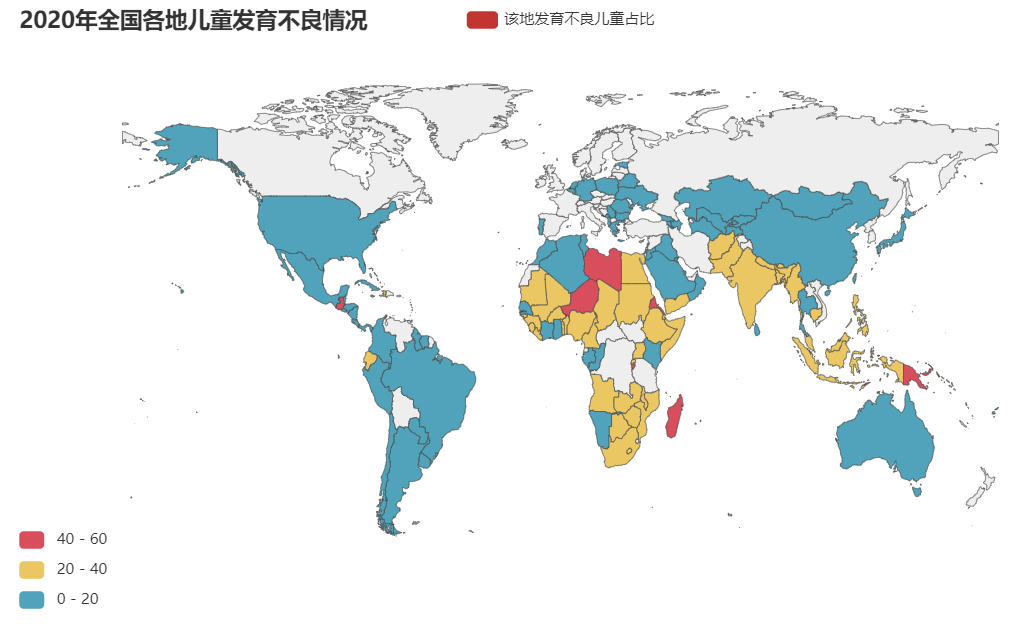
1. 绘制地理图并进行分析；
2. 绘制柱状图并进行分析

**4-1.绘制地理图进行分析**

读取预处理好的数据，绘制可视化图，将数值的大小分为三段，代表不同程度的营养不良情况。如下图所示。

|  |
| --- |
| df\_sheet1\_Point\_Estimate = df\_sheet1[df\_sheet1['Estimate'].isin(['Point Estimate'])]  Point\_Estimate = df\_sheet1\_Point\_Estimate.iloc[:,25:26]  Point\_Estimate=Point\_Estimate.astype(float)  df\_2020=pd.concat([country\_and\_areas,Point\_Estimate],axis=1)  import pandas as pd  from pyecharts import options as opts  from pyecharts.charts import Map  map1 = (  Map()  .add("该地发育不良儿童占比", [list(z) for z in zip(df\_2020['Country and areas'], df\_2020['2020'])], "world",is\_map\_symbol\_show=False)  .set\_series\_opts(label\_opts=opts.LabelOpts(is\_show=False))  .set\_global\_opts(  title\_opts=opts.TitleOpts(title="2020年全国各地儿童发育不良情况"),  visualmap\_opts=opts.VisualMapOpts(max\_=60,min\_=0,is\_piecewise=True,  pieces=[{"min": 0, "max": 20},{"min": 20, "max": 40},{"min": 40, "max": 60}])  )  )  map1.render\_notebook()  df\_sheet1\_Point\_Estimate = df\_sheet1[df\_sheet1['Estimate'].isin(['Point Estimate'])]  Point\_Estimate = df\_sheet1\_Point\_Estimate.iloc[:,5:6]  Point\_Estimate=Point\_Estimate.astype(float)  df\_2000=pd.concat([country\_and\_areas,Point\_Estimate],axis=1)  df\_2000.head()  import pandas as pd  from pyecharts import options as opts  from pyecharts.charts import Map  map2 = (  Map()  .add("该地发育不良儿童占比", [list(z) for z in zip(df\_2000['Country and areas'], df\_2000['2000'])], "world",is\_map\_symbol\_show=False)  .set\_series\_opts(label\_opts=opts.LabelOpts(is\_show=False))  .set\_global\_opts(  title\_opts=opts.TitleOpts(title="2000年全国各地儿童发育不良情况"),  visualmap\_opts=opts.VisualMapOpts(max\_=60,min\_=0,is\_piecewise=True,  pieces=[{"min": 0, "max": 20},{"min": 20, "max": 40},{"min": 40, "max": 60}])  )  )  map2.render\_notebook() |



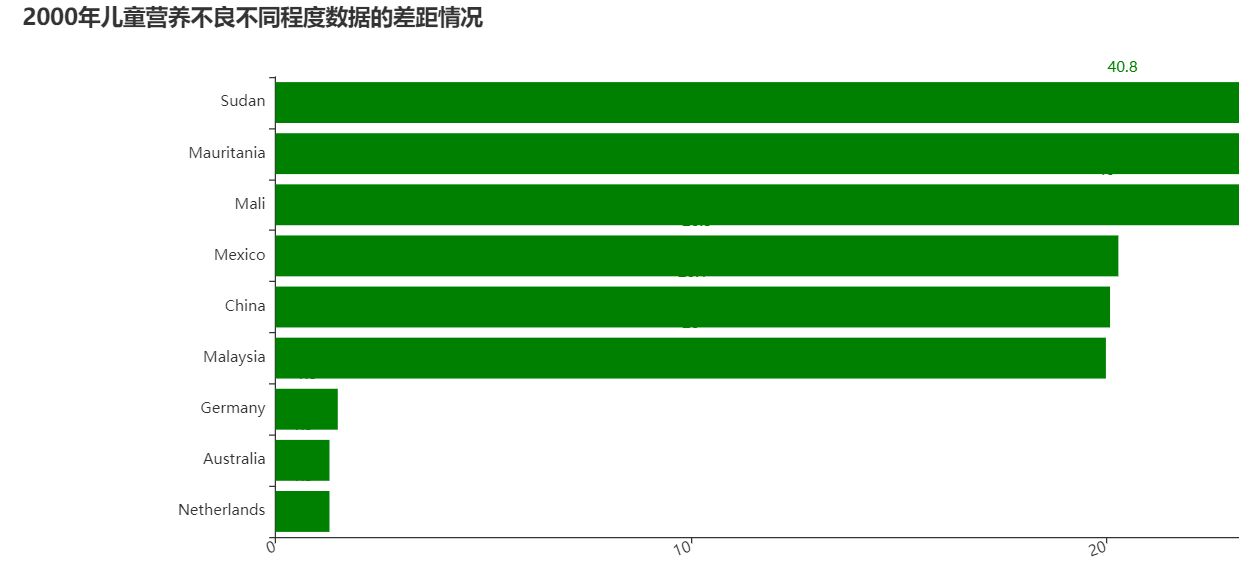


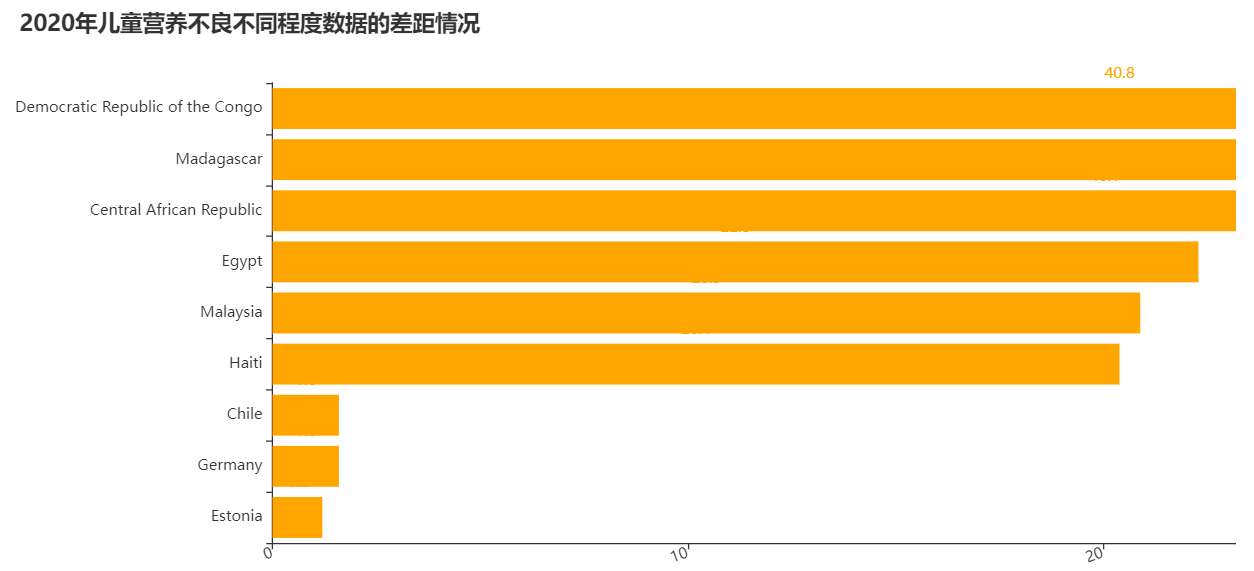
根据以上图片可以看出，相较于2000年，2020年更多的国家处于0%-20%此阶段，许多原先为40%-60%阶段的国家已进入20%-40%的阶段，仍有极少数的国家处于40%-60%这一阶段。

**4-2.绘制柱状图并进行分析**

读取数据并进行从小到大的排序，分别取出每个阶段的前三个国家，绘制柱状图。如下图所示。

|  |
| --- |
| df\_2020 = df\_2020.sort\_values('2020',ascending=True)  first\_2020 = df\_2020[:3]  second\_2020 = df\_2020[94:97]  third\_2020 = df\_2020[145:148]  first\_2020c = first\_2020['Country and areas']  first\_2020c = first\_2020c.values.tolist()  second\_2020c = second\_2020['Country and areas']  second\_2020c = second\_2020c.values.tolist()  third\_2020c =third\_2020['Country and areas']  third\_2020c = third\_2020c.values.tolist()  x = first\_2020c + second\_2020c + third\_2020c  first\_2020f = first\_2020['2020']  first\_2020f = first\_2020f.values.tolist()  second\_2020f = second\_2020['2020']  second\_2020f = second\_2020f.values.tolist()  third\_2020f = third\_2020['2020']  third\_2020f = third\_2020f.values.tolist()  y = first\_2020f + second\_2020f + third\_2020f  bar2 = (  Bar(init\_opts=opts.InitOpts(width="2100px",  height="500px"))  .add\_xaxis(x)  .add\_yaxis("",y,color='orange')  .set\_series\_opts(label\_opts=opts.LabelOpts(is\_show=True))  .reversal\_axis()  .set\_global\_opts(title\_opts=opts.TitleOpts(title="2020年儿童营养不良不同程度数据的差距情况"),  xaxis\_opts=opts.AxisOpts(axislabel\_opts={"rotate":20}),  legend\_opts=opts.LegendOpts(pos\_top=30,pos\_left=10))  )  bar2.render\_notebook()  df\_2000 = df\_2000.sort\_values('2000',ascending=True)  first\_2000 = df\_2000[:3]  second\_2000 = df\_2000[59:62]  third\_2000 = df\_2000[115:118]  first\_2000c = first\_2000['Country and areas']  first\_2000c = first\_2000c.values.tolist()  second\_2000c = second\_2000['Country and areas']  second\_2000c = second\_2000c.values.tolist()  third\_2000c =third\_2000['Country and areas']  third\_2000c = third\_2000c.values.tolist()  x = first\_2000c + second\_2000c + third\_2000c  first\_2000f = first\_2000['2000']  first\_2000f = first\_2000f.values.tolist()  second\_2000f = second\_2000['2000']  second\_2000f = second\_2000f.values.tolist()  third\_2000f = third\_2000['2000']  third\_2000f = third\_2000f.values.tolist()  y = first\_2000f + second\_2000f + third\_2000f  bar3 = (  Bar(init\_opts=opts.InitOpts(width="2100px",  height="500px"))  .add\_xaxis(x)  .add\_yaxis("",y,color='green')  .set\_series\_opts(label\_opts=opts.LabelOpts(is\_show=True))  .reversal\_axis()  .set\_global\_opts(title\_opts=opts.TitleOpts(title="2000年儿童营养不良不同程度数据的差距情况"),  xaxis\_opts=opts.AxisOpts(axislabel\_opts={"rotate":20}),  legend\_opts=opts.LegendOpts(pos\_top=30,pos\_left=10))  )  bar3.render\_notebook() |





由上图可看出，经过20年的发展，各个阶段前三个国家大多数早已改变，但也有极少数国家如

’Germany’原先处于0%-20%阶段的第三名，2020年已处于第二名。

‘Malaysia’原先处于20%-40%阶段的第一名，2020年跌到20-40阶段的第二名。

**5.儿童超重情况可视化分析**

**任务描述：**

对各国儿童超重的比例和人数进行划分，根据数值 的大小区分超重的程度，绘制图表，统计并对比 2000 年与 2020 年不同程度分布情况；绘制 地理图，分析不同国家儿童营养不良的情况；根据不同程度，抽取排名前 3 的国家数据，整 合数据分析儿童营养不良不同程度数据的差距情况；分析对比 2000 年与 2020 年各国儿童 超重的比例和人数变动情况。

**任务分析：**

对总体情况可视化可以分为以下步骤。

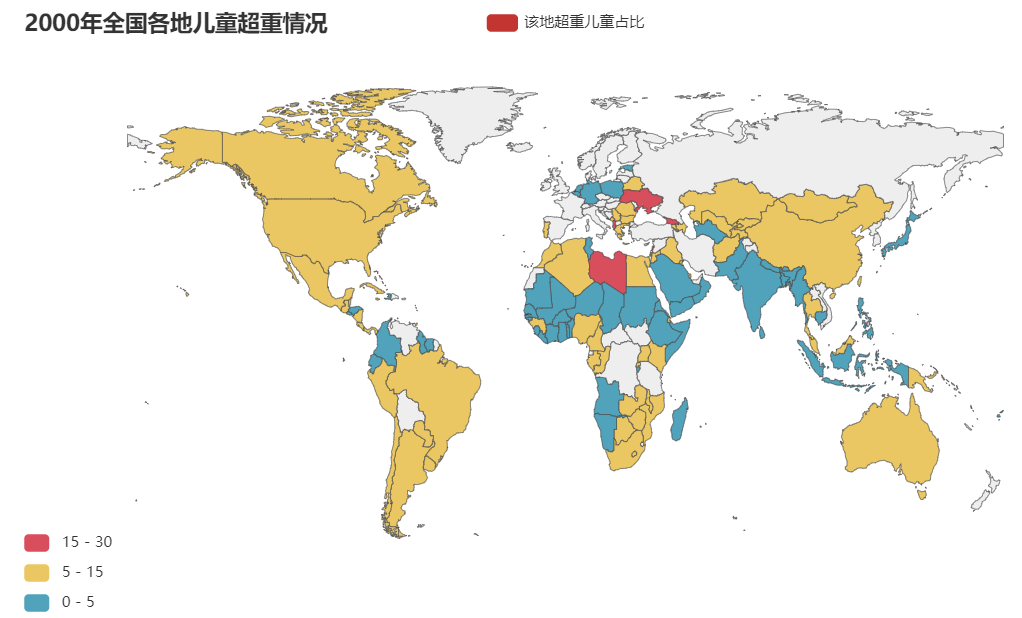
（1）绘制地理图并进行分析；

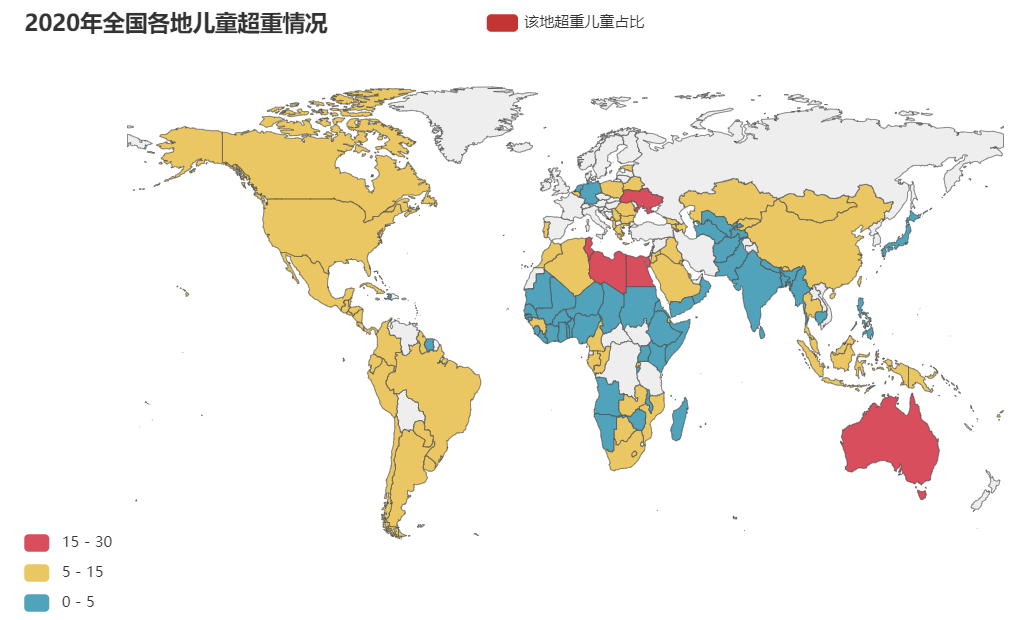
（2）绘制柱状图并进行分析

**5-1.绘制地理图并进行分析**

读取预处理好的数据，绘制可视化图，将数值的大小分为三段，代表不同程度的超重情况。如下图所示。

|  |
| --- |
| 提取数据  df\_sheet3\_Point\_Estimate = df\_sheet3[df\_sheet3['Estimate'].isin(['Point Estimate'])]  Point\_Estimate3 = df\_sheet3\_Point\_Estimate.iloc[:,5:6]  Point\_Estimate3 = Point\_Estimate3.astype(float)  df\_2000 = pd.concat([country\_and\_areas3,Point\_Estimate3],axis=1)  import pandas as pd  from pyecharts import options as opts  from pyecharts.charts import Map  画2000超重图  map3 = (  Map()  .add("该地超重儿童占比", [list(z) for z in zip(df\_2000['Country and areas'], df\_2000['2000'])], "world",is\_map\_symbol\_show=False)  .set\_series\_opts(label\_opts=opts.LabelOpts(is\_show=False))  .set\_global\_opts(  title\_opts=opts.TitleOpts(title="2000年全国各地儿童超重情况"),  visualmap\_opts=opts.VisualMapOpts(max\_=60,min\_=0,is\_piecewise=True,  pieces=[{"min": 0, "max": 5},{"min": 5, "max": 15},{"min": 15, "max": 30}])  )  )  map3.render\_notebook()  #提取数据  df\_sheet3\_Point\_Estimate = df\_sheet3[df\_sheet3['Estimate'].isin(['Point Estimate'])]  Point\_Estimate3 = df\_sheet3\_Point\_Estimate.iloc[:,25:26]  Point\_Estimate3 = Point\_Estimate3.astype(float)  df\_2020 = pd.concat([country\_and\_areas3,Point\_Estimate3],axis=1)  import pandas as pd  from pyecharts import options as opts  from pyecharts.charts import Map  画2020超重图  map4 = (  Map()  .add("该地超重儿童占比", [list(z) for z in zip(df\_2020['Country and areas'], df\_2020['2020'])], "world",is\_map\_symbol\_show=False)  .set\_series\_opts(label\_opts=opts.LabelOpts(is\_show=False))  .set\_global\_opts(  title\_opts=opts.TitleOpts(title="2020年全国各地儿童超重情况"),  visualmap\_opts=opts.VisualMapOpts(max\_=60,min\_=0,is\_piecewise=True,  pieces=[{"min": 0, "max": 5},{"min": 5, "max": 15},{"min": 15, "max": 30}])  )  )  map4.render\_notebook() |



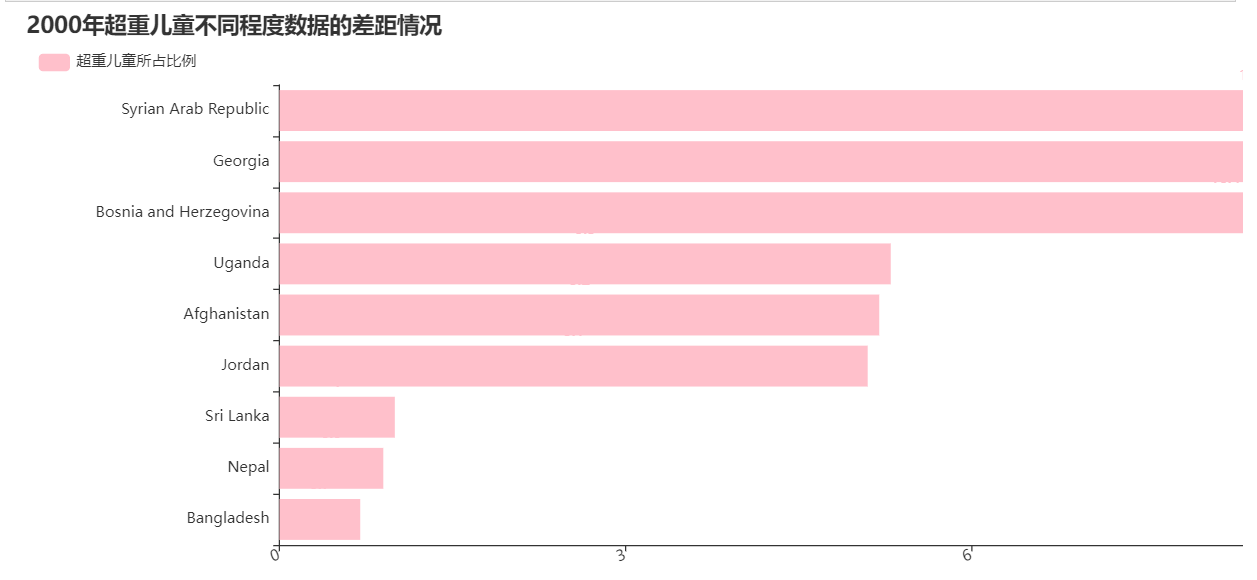


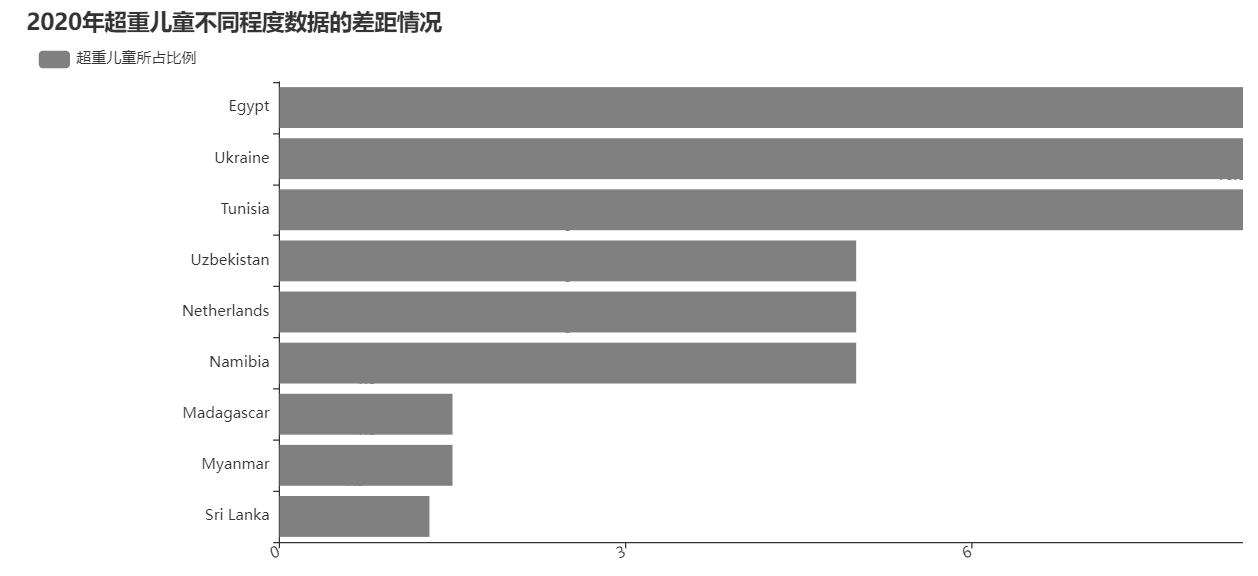
对比2000年及2020年全国各地儿童超重情况可以看出，大部分国家儿童超重情况都处于5%-15%阶段并且经过20年仍然不变，少部分国家儿童超重占比变大，甚至出现从0%-5%阶段进入5%-15%阶段，从5%-15%阶段进入15%-30%阶段。

**5.2绘制柱状图并进行分析**

读取数据并进行从小到大的排序，分别取出每个阶段的前三个国家，绘制柱状图。如下图所示。

|  |
| --- |
| df\_2000 = df\_2000.sort\_values('2000',ascending=True)  first\_2000 = df\_2000[:3]  second\_2000 = df\_2000[65:68]  third\_2000 = df\_2000[148:151]  first\_2000c = first\_2000['Country and areas']  first\_2000c = first\_2000c.values.tolist()  second\_2000c = second\_2000['Country and areas']  second\_2000c = second\_2000c.values.tolist()  third\_2000c =third\_2000['Country and areas']  third\_2000c = third\_2000c.values.tolist()  x = first\_2000c + second\_2000c + third\_2000c  first\_2000f = first\_2000['2000']  first\_2000f = first\_2000f.values.tolist()  second\_2000f = second\_2000['2000']  second\_2000f = second\_2000f.values.tolist()  third\_2000f = third\_2000['2000']  third\_2000f = third\_2000f.values.tolist()  y = first\_2000f + second\_2000f + third\_2000f  bar4 = (  Bar(init\_opts=opts.InitOpts(width="2100px",  height="500px"))  .add\_xaxis(x)  .add\_yaxis("超重儿童所占比例",y,color='pink')  .set\_series\_opts(label\_opts=opts.LabelOpts(is\_show=True))  .reversal\_axis()  .set\_global\_opts(title\_opts=opts.TitleOpts(title="2000年超重儿童不同程度数据的差距情况"),  xaxis\_opts=opts.AxisOpts(axislabel\_opts={"rotate":20}),  legend\_opts=opts.LegendOpts(pos\_top=30,pos\_left=10))  )  bar4.render\_notebook()  df\_2020 = df\_2020.sort\_values('2020',ascending=True)  first\_2020 = df\_2020[:3]  second\_2020 = df\_2020[57:60]  third\_2020 = df\_2020[148:151]  first\_2020c = first\_2020['Country and areas']  first\_2020c = first\_2020c.values.tolist()  second\_2020c = second\_2020['Country and areas']  second\_2020c = second\_2020c.values.tolist()  third\_2020c =third\_2020['Country and areas']  third\_2020c = third\_2020c.values.tolist()  x = first\_2020c + second\_2020c + third\_2020c  first\_2020f = first\_2020['2020']  first\_2020f = first\_2020f.values.tolist()  second\_2020f = second\_2020['2020']  second\_2020f = second\_2020f.values.tolist()  third\_2020f = third\_2020['2020']  third\_2020f = third\_2020f.values.tolist()  y = first\_2020f + second\_2020f + third\_2020f  bar5 = (  Bar(init\_opts=opts.InitOpts(width="2100px",  height="500px"))  .add\_xaxis(x)  .add\_yaxis("超重儿童所占比例",y,color='grey')  .set\_series\_opts(label\_opts=opts.LabelOpts(is\_show=True))  .reversal\_axis()  .set\_global\_opts(title\_opts=opts.TitleOpts(title="2020年超重儿童不同程度数据的差距情况"),  xaxis\_opts=opts.AxisOpts(axislabel\_opts={"rotate":20}),  legend\_opts=opts.LegendOpts(pos\_top=30,pos\_left=10))  )  bar5.render\_notebook() |





对比2000年与2020年三个阶段的儿童超重占比前三名发现，各个阶段前三个国家大多数早已改变，但‘Sri Lanka’从原先的在全世界中儿童超重占比最小第三名转变为第一名。

**6.可视化分析结果**

通过对儿童身体健康情况数据进行分析，可视化后可以得出以下结论

1. 随着经济发展，科技水平的提供，全世界范围内营养不良人数逐渐减少
2. 世界发育不良儿童占比减低，但处于非洲地区的国家仍然面临着儿童发育不良占比增加的现象
3. 大部分发育不良占比少的国家其儿童超重占比呈现增长趋势，而发育不良占比大的国家则儿童超重占比减低，其速度小于世界儿童超重增长趋势，形成两极分化的现象
4. 发达国家，大部分发展中国家儿童健康情况良好，落后国家情况较差。
5. 在物质水平提高的同时更要注重身体锻炼，要珍惜来之不易的食物，节约粮食。