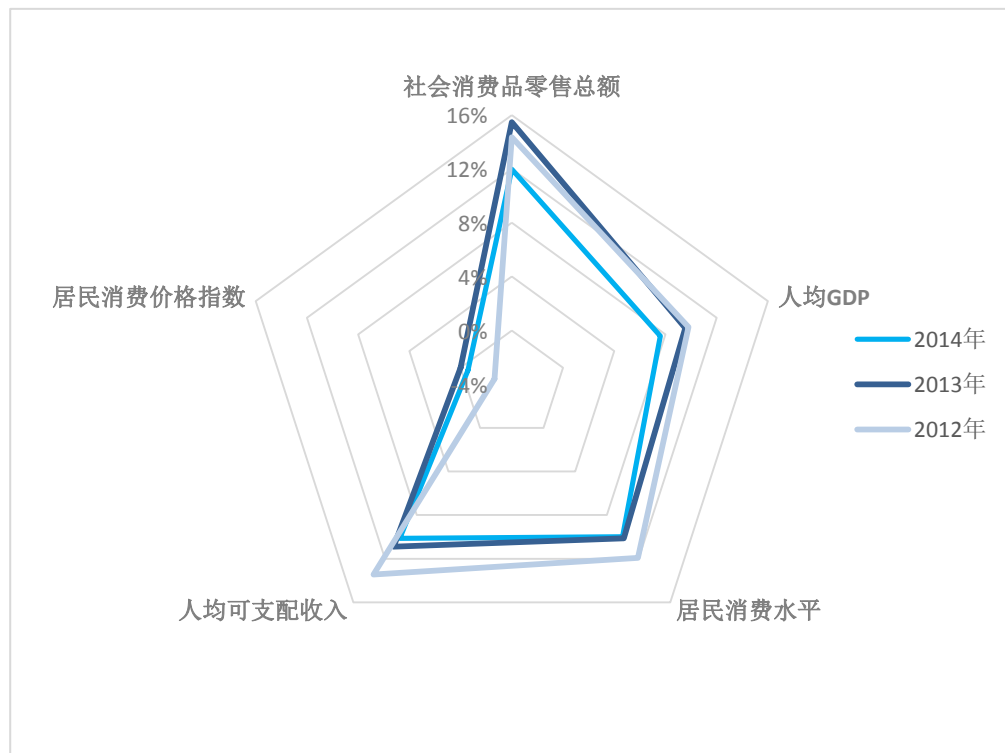


练习题1：图解中国人的生活水平变化



根据统计局的数据，显示近三年人均GDP、居民消费水平、人均可支配收入同比增长率，均呈现下降趋势，增速放缓。2014年社会消费品零售总额同比增长率，也低于前两年。居民消费价格指数同比增长率，在近三年中居中。

中国人的生活水平变化，不一定只通过这些指标来参考，还有很多数据可以看，比如说近几年出国游大热，海外购物那更是把别国奶粉都买光了。另一方面随着贫富差距的增大，很多人被平均值拉高了；再者很多数据是抽样数据，本身存在一定误差。

练习2：为什么要双盲实验，如何应用到具体某个场景

双盲实验用于防止研究结果被安慰剂效应(placebo effect)或者观察者偏爱(observer bias)影响。在实验中使一个人盲就是不告知他实验过程的信息。按照使参与者盲的程度，这种实验被分为单盲(Single-blind)、双盲(double-blind)、三盲(triple-blind)实验。

练习3：研究你关心的某一词组的变化



练习4：思考为什么明明会出现simpson's paradox，对于其结果如何利用

当人们尝试探究两种变量(比如新生录取率与性别)是否具有相关性的时候，会分别对之进行分组研究。然而，在分组比较中都占优势的一方，在总评中有时反而是失势的一方，即辛普森悖论

为了避免出现辛普森悖论，对于分组的数据，要考虑增加权重，以一定的系数去除分组基数差异造成的影响；另一方面，同时必须了解所设情境是否存在其他潜在因素的影响，需要综合考虑的

练习5：贝叶斯定理

根据贝叶斯定理的公式，
$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{P(A)} = \frac{P(B_i)P(A|B_i)}{\sum P(AB_j)} = \frac{P(B_i)P(A|B_i)}{\sum P(A|B_j)P(B_j)}$$

在此事件中，A代表检验结果呈阳性， $P(B_1)$ 代表患病， $P(B_2)$ 代表不患病
某种疾病的发病率是0.001,即 $P(B_1)=0.1\%$ ；不发病率，即 $P(B_2) = 99.9\%$
在患者确实得病的情况下，它有99%的可能呈现阳性，即 $P(A|B_1) = 99\%$
在患者没有得病的情况下，它有5%的可能呈现阳性，即 $P(A|B_2) = 5\%$
现有一病人的检验结果为阳性，请问他确实得病的可能性有多大，即 $P(B_1|A)$

$$P(B_1|A) = \frac{P(B_1)P(A|B_1)}{P(A)} = \frac{P(B_1)P(A|B_1)}{\sum P(AB_j)} = \frac{P(B_1)P(A|B_1)}{\sum P(A|B_j)P(B_j)} = 1.943463\%$$

