



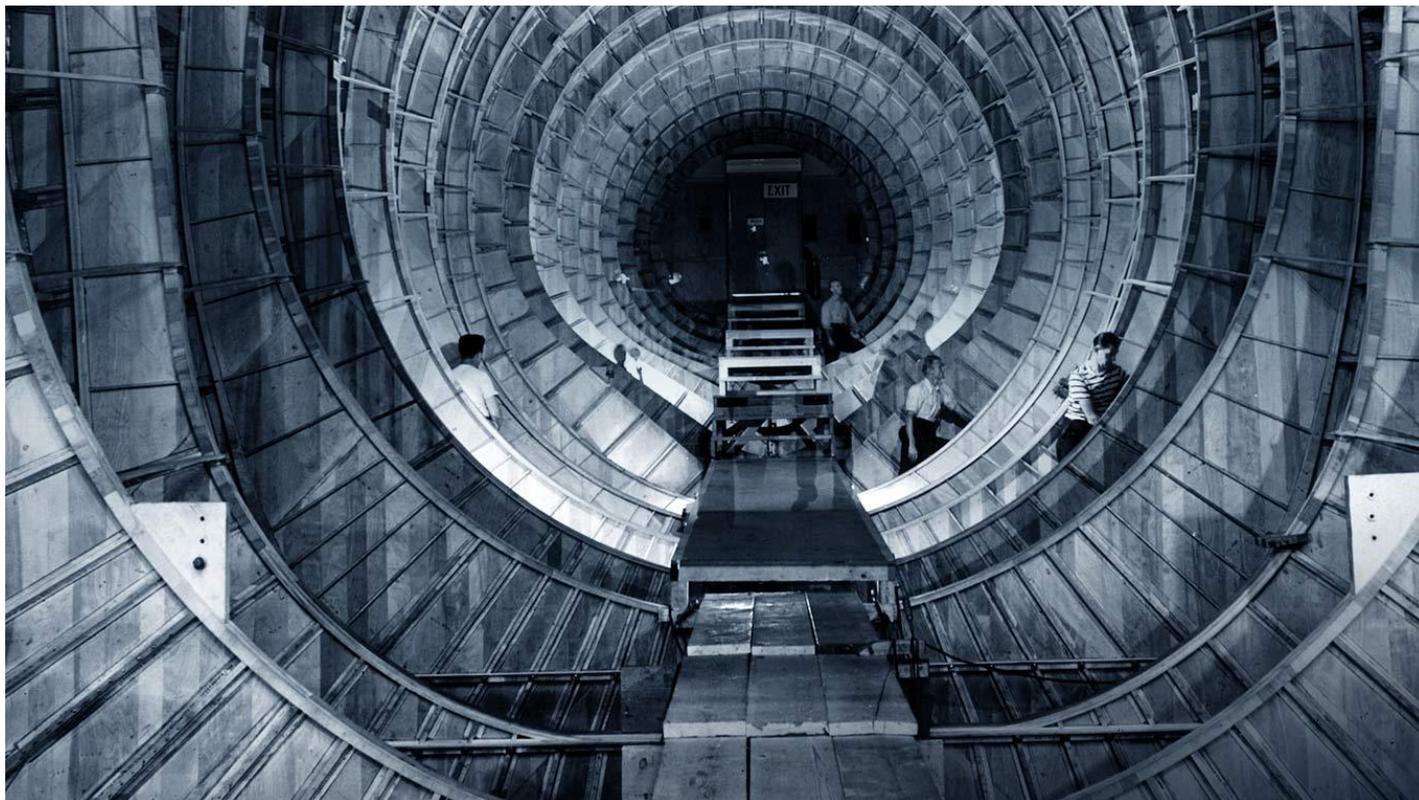
广西省少数民族分布及发展状况

王露、付颖、喻易浩佳



目录

CONTENTS



第一部分 | 背景

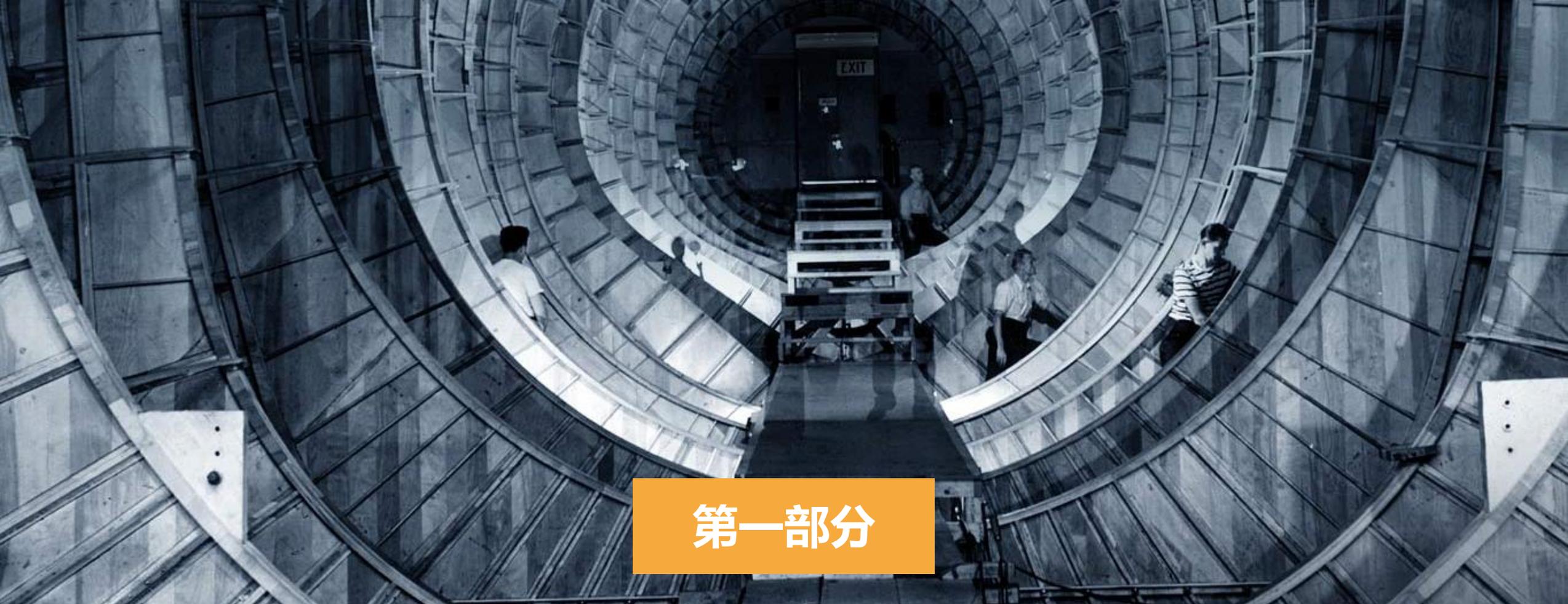
第二部分 | 聚类指标选择

第三部分 | 数据初步处理

第四部分 | 聚类分析

第五部分 | 结论

第六部分 | 附录



第一部分

背景

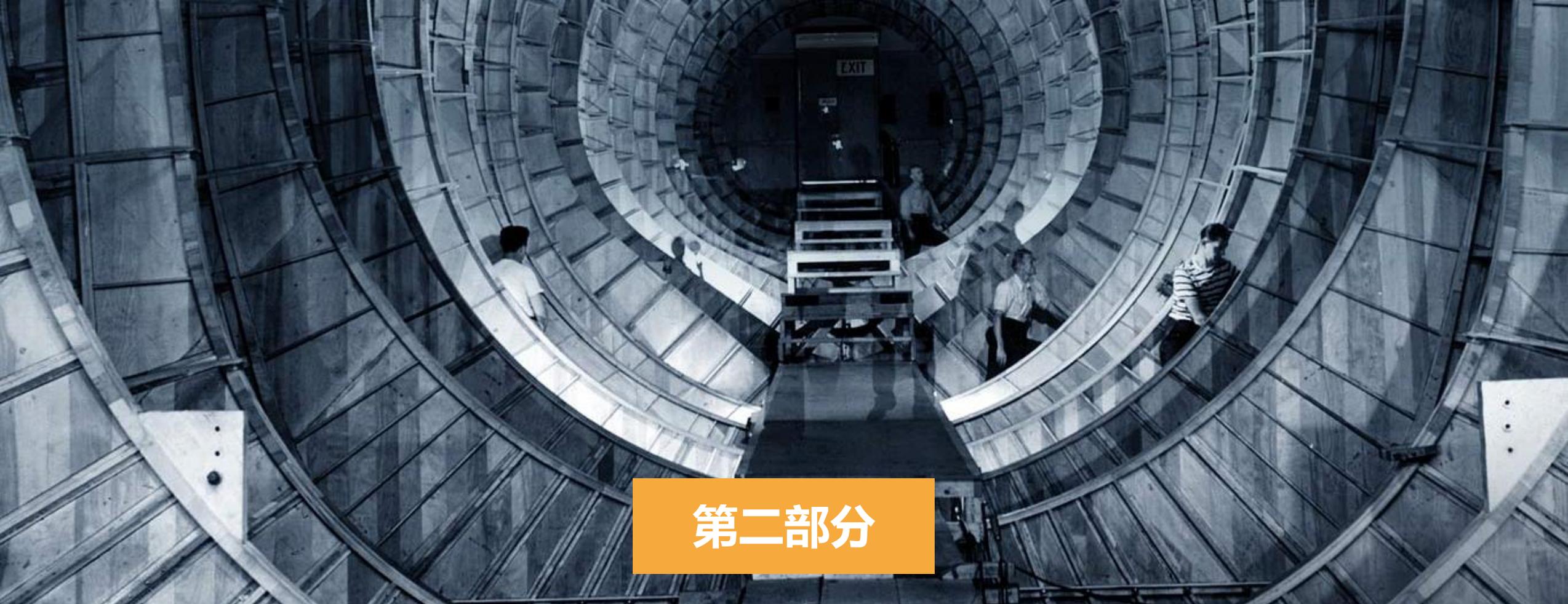


广西壮族自治区

广西省为我国五个少数民族自治区之一，拥有我国最多的少数民族人口数量，地方语言风俗丰富。

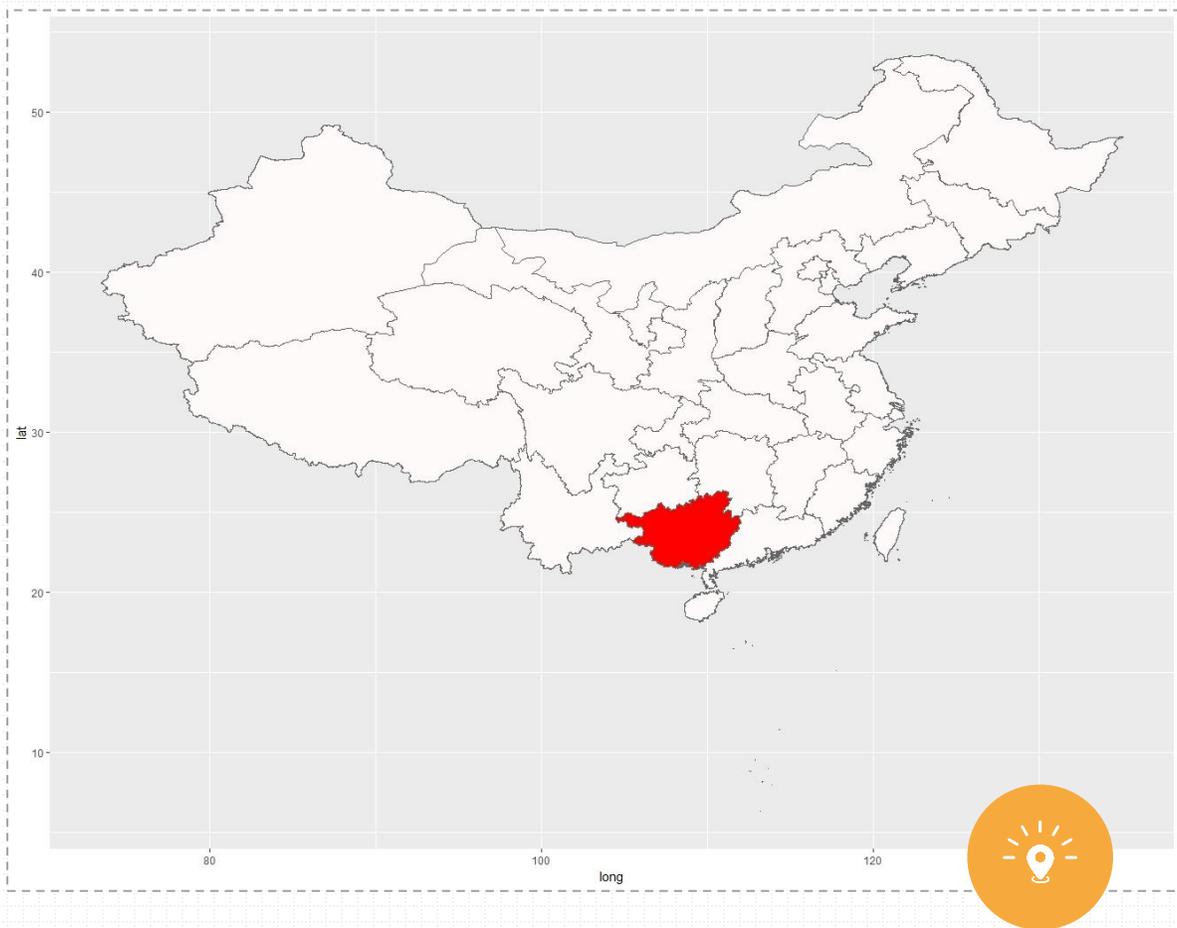
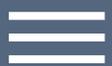
研究意义

了解广西省少数民族分布及发展状态有利于少数民族相关政策制定、旅游业开发。



第二部分

聚类指标选择



少数民族个数

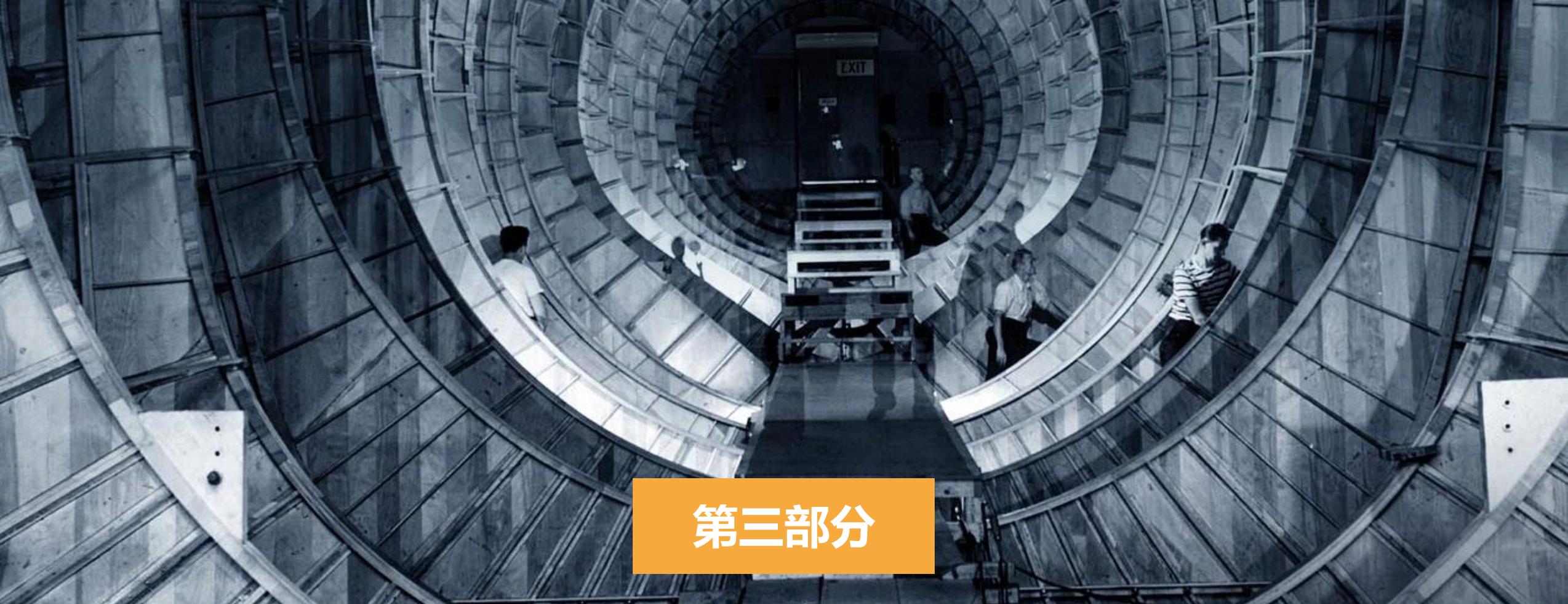
以市为单位的民族个数能反应当地人口组成状况。复杂程度、多元化程度。

GDP

GDP是一个城市发展状况的反映。

常住人口

常住人口较户籍人口更为真实的反映了城市人口聚集情况。

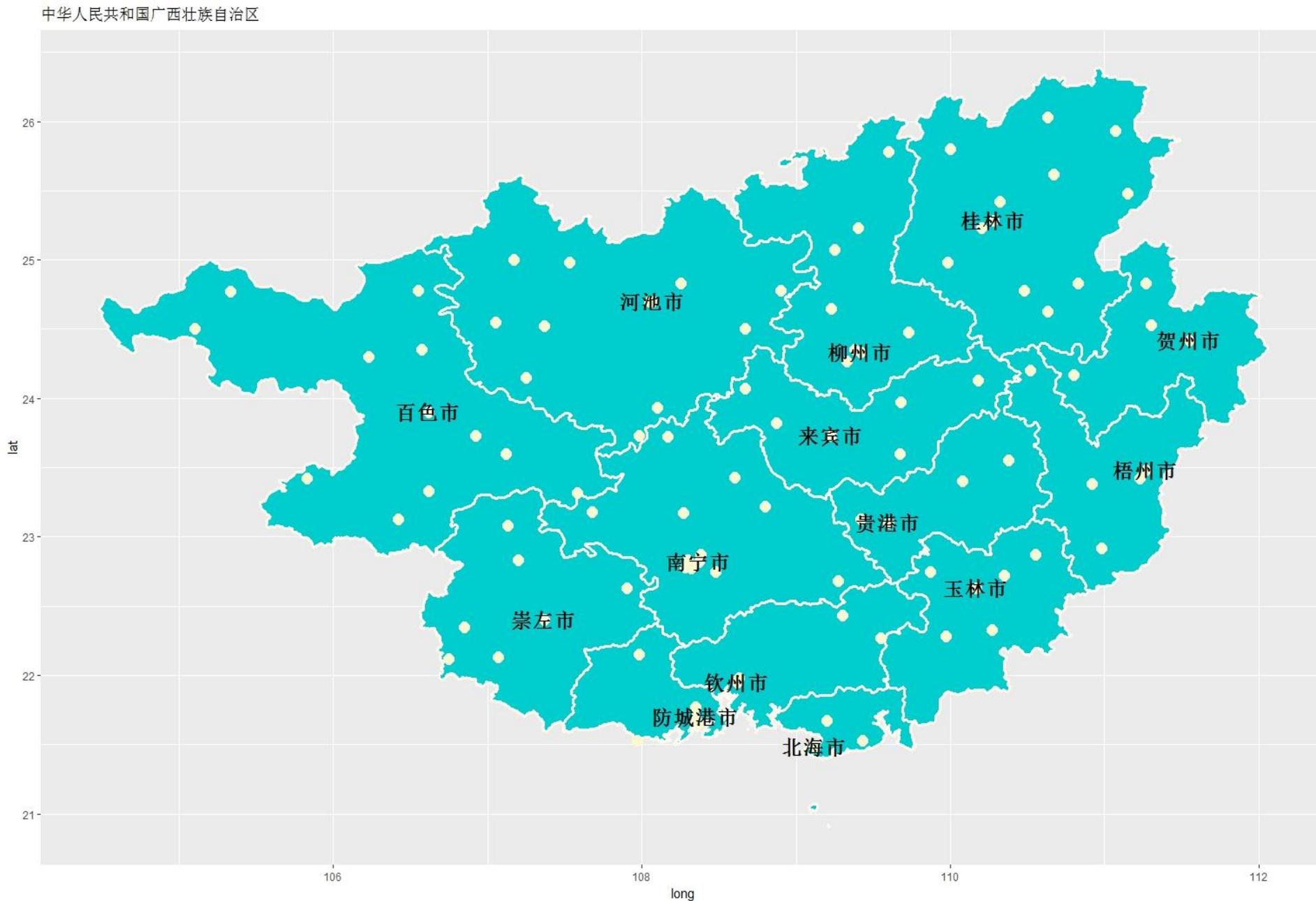


第三部分

数据初步处理

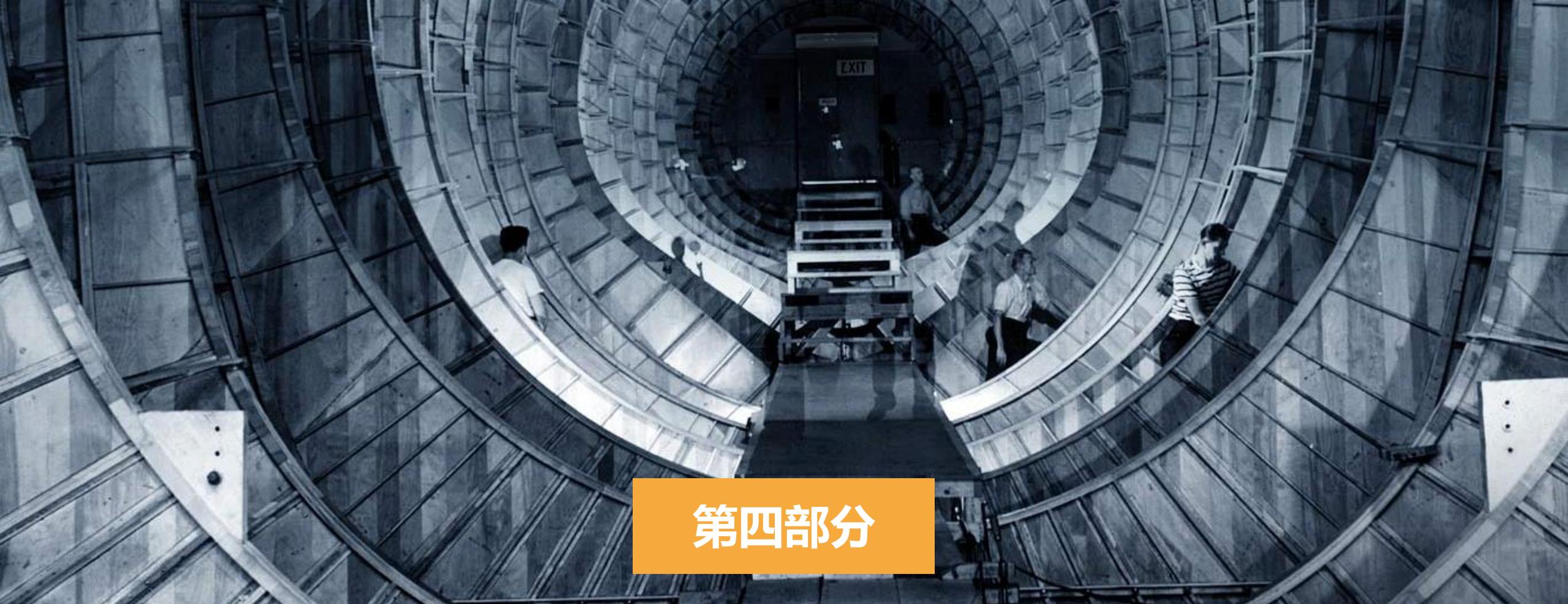


第三部分 | 广西壮族自治区14市级行政区





地级市	少数民族数	GDP	常住人口
南宁市	2	341009	6980100
柳州市	6	229862	3992700
桂林市	4	194297	4961600
梧州市	0	107859	2999400
北海市	0	89208	1625700
防城港市	1	62072	918400
贵港市	0	86520	4293700
玉林市	0	144613	5707200
钦州市	0	94442	3209300
崇左市	1	68282	2054500
来宾市	2	55770	2182000
贺州市	1	46811	2025900
百色市	5	98035	3596700
河池市	6	61803	3476800

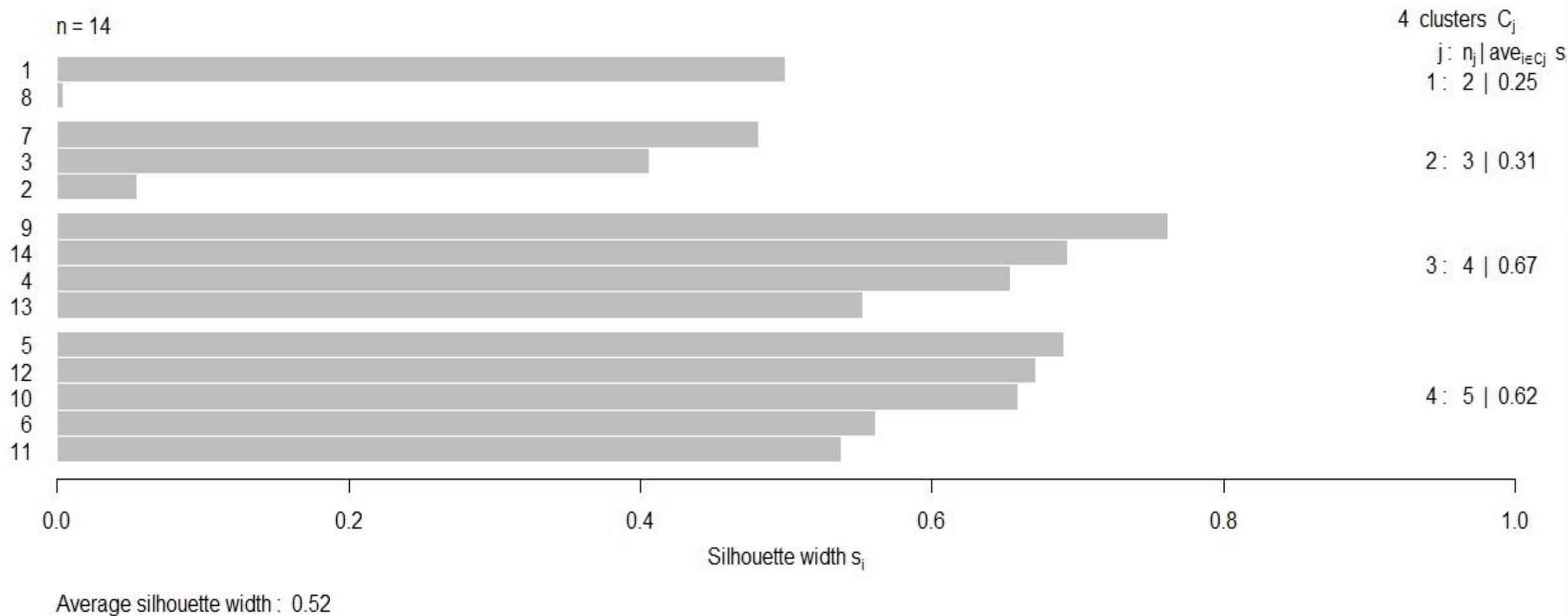


第四部分

聚类分析

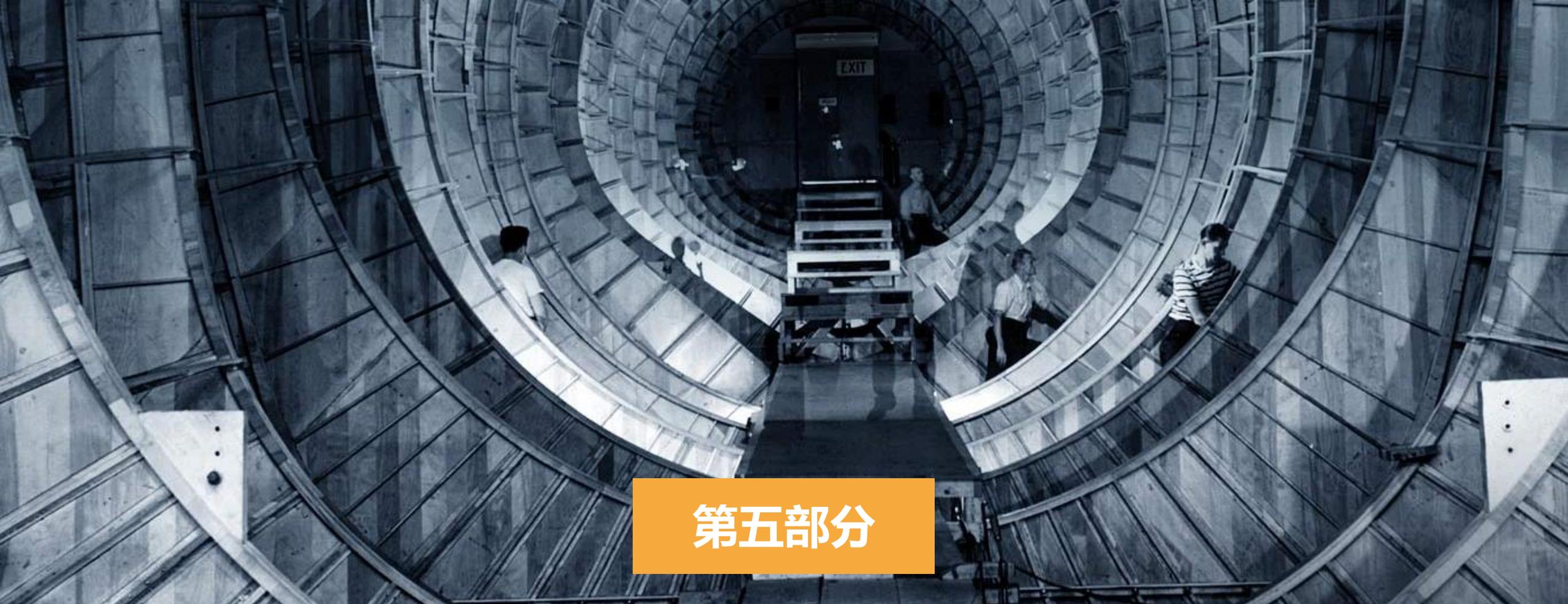


第四部分 | 广西壮族自治区数据聚类分析结果



1.南宁 2.柳州 3.桂林 4.梧州 5.北海 6.防城港 7.贵港 8.玉林 9.钦州 10.崇左 11.来宾 12.贺州 13.百色 14.河池





第五部分

结论

结论一

汉族要带动少数民族发展

总体来看，常住人口多，汉族所占比例较大的城市，GDP较高，城市发展状况较好。因此，汉族要带动少数民族发展。

结论二

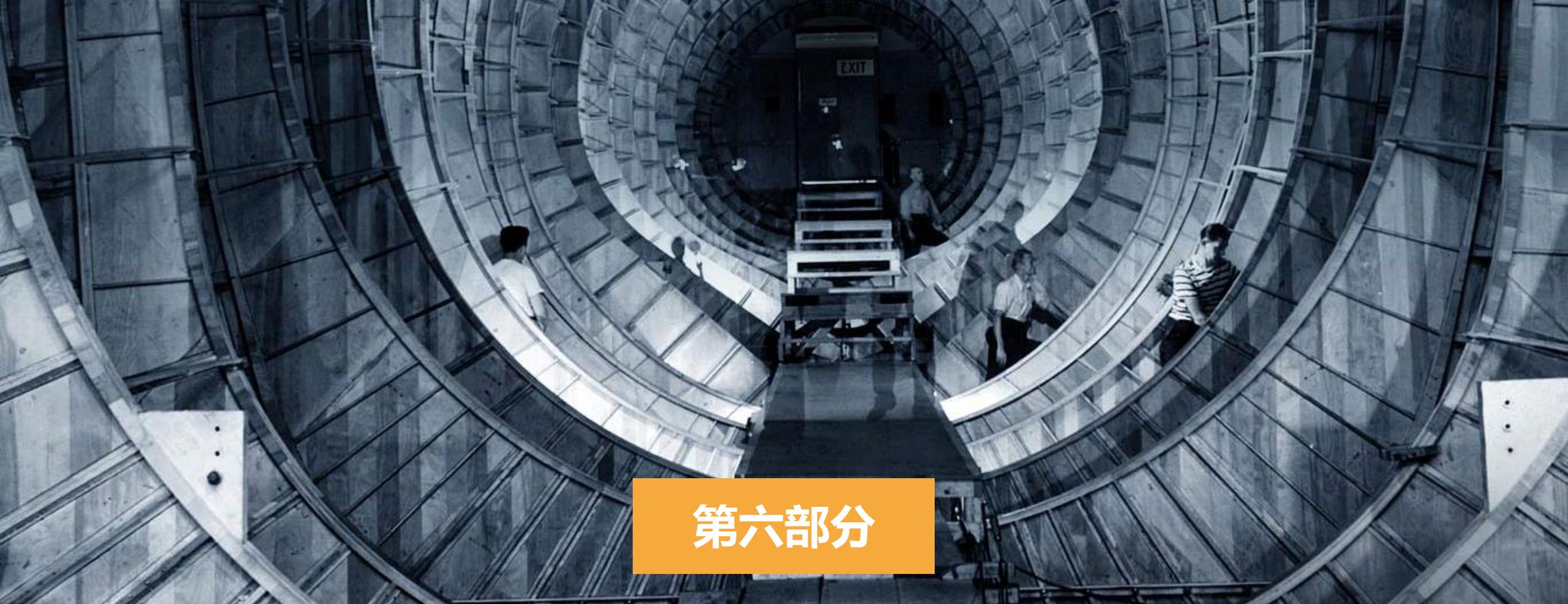
少数民族要充分发挥自己的优势

二类城市的特点主要是人口中等，GDP较高。其中桂林、柳州两个城市聚集了较多的少数民族，城市发展状况较好，和各自城市的旅游业发展有一定的关系。

结论三

少数民族要相互学习，共同发展

三类城市中，百色和河池少数民族数量多，常住人口多，但城市发展较慢。位于山区的地理位置有一定影响。可以向桂林和柳州学习，对本地资源进行有效开发。



第六部分

附录



```
setwd("C:/Users/WangLu/Desktop")
install.packages("maptools")
install.packages("ggplot2")
install.packages("plyr")
library(maptools)
#画出广西壮族自治区位于中国的位置
x=readShapePoly('bou2_4p.shp')
w=x@data
xs=data.frame(w,id=seq(0:924)-1)
library(ggplot2)
china_map<-fortify(x)
library(plyr)
china_mapdata<-join(china_map, xs, type = "full")
ggplot(china_mapdata, aes(x = long, y = lat, group = group,fill=NAME=="广西壮族自治区"))+
  geom_polygon()+
  geom_path(colour = "grey40")+
  scale_fill_manual(values=c("snow","red"),guide=FALSE)
```



```
#读入广西壮族自治区的市县位置
y=readShapePoly('CHN_adm2.shp')
k=subset(y,y$NAME_1=='Guangxi')
guangxi=k@data
gxs=data.frame(guangxi,id=seq(0:13)-1)
guangximap=fortify(k)
guangximapdata=join(guangximap,gxs,type="full")
#读入表格
install.packages("XML")
library(XML)
webpage = "http://blog.csdn.net/svrsimon/article/details/8255051"
tables = readHTMLTable(webpage, stringsAsFactors = FALSE, header=T)
tables=as.data.frame(tables[1])
#找出广西壮族自治区数据，区分市和区县，保存表格
gxtables=subset(tables,tables$NULL.省份=="广西壮族自治区")
gxtables$是否为地级市=(gxtables$NULL.地市==gxtables$NULL.区县)
write.csv(gxtables,"广西壮族自治区市区县位置.csv")
gxcitytables=subset(gxtables,gxtables$是否为地级市==TRUE)
```



```
#添加市县区标签，画出广西地图
gxtables$NULL.经度=as.numeric(gxtables$NULL.经度)
gxtables$NULL.纬度=as.numeric(gxtables$NULL.纬度)
ggplot(guangximapdata, aes(x = long, y = lat, group = group, fill=NAME_2))+
  ggtitle("中华人民共和国广西壮族自治区")+
  geom_polygon(fill="cyan3")+
  geom_path(colour = "floralwhite",size=1.3)+
  scale_fill_manual(values=colours(),guide=FALSE)+
  annotate("pointrange",x=gxtables$NULL.经度, y=gxtables$NULL.纬度, ymin=gxtables$NULL.
  纬度,
    colour="lightgoldenrodyellow", size=1)+
  annotate("text",x=as.numeric(gxcitytables$NULL.经度), y=as.numeric(gxcitytables$NULL.
  纬度), label=gxcitytables$NULL.
  区县, family="serif",
    fontface="bold", colour="black", size=6)
#根据民族统计情况，聚类上色
clusterdata=read.csv("guangxi.csv")
install.packages("cluster")
require(cluster)
keep.cols<-which(names(clusterdata)%in% c("少数民族数","GDP","常住人口"))
```



```
gxPam <- pam(x=clusterdata[keep.cols], k=4, keep.diss=TRUE,
            keep.data=TRUE)
gxPam$medoids
plot(gxPam, which.plots=2, main="")
require(plyr)
k@data$NAME_2=as.character(revalue(k@data$NAME_2,
                                replace = c(Baise="百色市",Beihai="北海市",Chongzuo="崇左市",Fangchenggang="防城港市",
                                             Guigang="贵港市",Guilin="桂林市",Hechi="河池市",Hezhou="贺州市",
                                             Laibin="来宾市",Liuzhou="柳州市",Nanning="南宁市",Qinzhou="钦州市",
                                             Wuzhou="梧州市",Yulin="玉林市"))))

k@data$id<-rownames(k@data)
require(ggplot2)
install.packages("rgeos")
install.packages("broom")
require(rgeos)
require(broom)
clusmap=fortify(k)
clusmapdata=join(clusmap,k@data[c("id","NAME_2")],by="id")
```



第六部分 | 代码

#生成含有聚类结果的表格

```
clusterMemebership=data.frame(NAME_2=c(as.character(clusterdata$地级市)),  
                               Cluster=factor(gxPam$clustering),  
                               stringsAsFactors = FALSE)
```

```
clusmapdata<-join(clusmapdata, clusterMemebership, by="NAME_2")
```

#绘制地图

```
ggplot()+ geom_polygon(data=clusmapdata, aes(x=long, y=lat, group=group, fill=Cluster, color=Cluster)) +  
  labs(x=NULL,y=NULL) + coord_equal() +  
  annotate("pointrange",x=gxtables$NULL.经度, y=gxtables$NULL.纬度, ymin=gxtables$NULL.纬度,ymax=gxtables$NULL.  
  纬度,  
         colour="lightgoldenrodyellow", size=0.6)+  
  annotate("text",x=as.numeric(gxcitytables$NULL.经度), y=as.numeric(gxcitytables$NULL.纬度), label=gxcitytables$NULL.  
  区县, family="serif",  
         fontface="bold", colour="black", size=6)+  
theme(panel.grid.major = element_blank(),  
      panel.grid.minor = element_blank(),  
      axis.text.x = element_blank(),  
      axis.text.y = element_blank(),  
      axis.ticks = element_blank(), panel.background = element_blank())
```



THANK YOU!

