R语言在CGED-Q JSL中的运用

陈俊
香港科技大学社会科学部
2025年5月(V2.1.1)

Eight

制作GIS图像

01

导入GIS底图

8.1 制作GIS图像
8.1.1 GIS图像简介

- 清晰直观
- 时空分布特征
- 主流GIS分析软件
  - ArcGIS
  - QGIS
- RStudio ggplot2
- 需要空间地理数据作为GIS底图

8.1.2 CHGIS (The China Historical Geographic Information System)

- 哈佛大学和复旦大学历史地理研究中心自2002年开始陆续公开6个版本的中国古代聚居区和行政区划数字地理信息数据库

- 开源
  - 供学术研究者免费试用
  - 可自行至CHGIS官网或复旦史地所下载
  - 时间截面和时间连续数据

- 时间截面数据1820、1911
  - 省、府、县（仅1911）的地理边界
  - 省、府、县、镇的地理点数据
  - 河流、湖泊、海岸线等水文数据（仅1820）
  - 太湖的区域性地理数据（仅1820）

- 与1900-1912年缙绅录数据库提供的的地理信息紧密相关

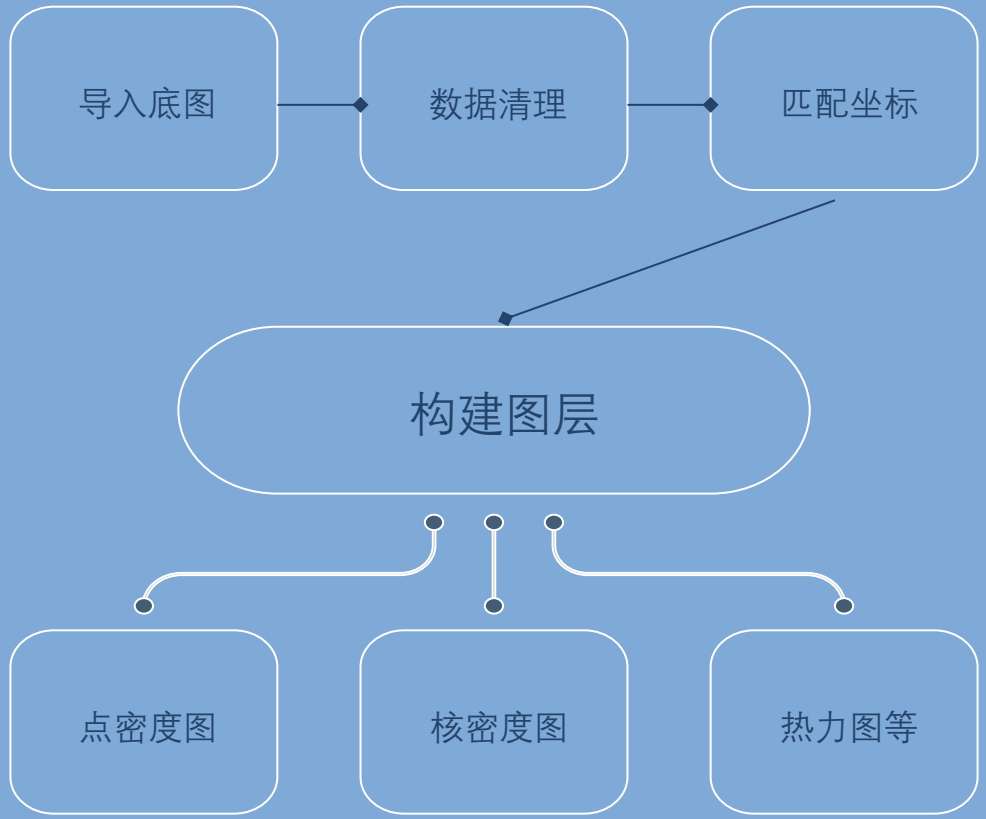
8.1.3 注意！！！地图使用规范

- 使用地图时必须遵守相关规范

- 如需要制作公开发表的地图
  - 请参照自然资源部的相关规范文件
  - https://www.gov.cn/gongbao/content/2023/content_5752310.htm
  - 或者使用自然资源部推出的标准地图服务
  - http://bzdt.ch.mnr.gov.cn

- 本教程考虑到简化代码、串联制图流程、直观展示分析研究结果的重要性，所以使用了CHGIS制作的1911年中国历史地图，本教程制作的所有地图仅供展示制图代码结果以及个人学习制图代码使用，严禁单独转发、盗用、误用本教程制作的历史地图。本教程是开源的公益教程，因此不具备实时监控能力，对于任何因不完整转发、盗用、误用本教程制作的所有历史地图而产生的问题，本教程概不负责，亦不负任何法律责任。

8.1.4 制作GIS图像步骤



```
# 安装地理信息处理包
install.packages(c(“sf”, "ggpointdensity”, "ggspatial”,  "ggnewscale”, "viridis”, "scales”))

# 读取包
library(“sf”, “ggplot2”, “dplyr”, “stringr”,  “ggpointdensity”, “ggspatial”, “ggnewscale”, “viridis”, “scales”)
```

8.1.5 导入数据

- 自行下载`CHGIS V6 shapefile`文件，导入1911年
  - 县级点、县级边界数据
  - 府级边界数据
  - 省级边界、省级点数据

- 另请读取.shp为后缀的文件，同时.shx以及.dbf等其他后缀文件需要和.shp在同一个文件夹之下，以便R调取

- 后缀
  - `pts`为地理点数据，可绘制坐标点，用于和各类数据进行匹配以赋予每条记录坐标
  - `pgn`为地理边界数据，可绘制边界，用于生成底图

- 两类文件作用不同，请明确区分两类文件的用法，否则会导致制图不成功

导入文件，请更换文件路径

```
chgis1911_county_pts <-
 “/Users/jc/Documents/6-分析-GIS底图/CHGIS/CHCIS1911county_gbk_shp_pts/v6_1911_cnty_pts_gbk.shp”
# 县级地理点数据

chgis1911_prov_pgn <-
 “/Users/jc/Documents/6-分析-GIS底图/CHGIS/CHCIS1911prov_gbk_shp_pgn/v6_1911_prov_pgn_gbk.shp”
# 省级边界数据

chgis1911_prov_pts <-
 “/Users/jc/Documents/6-分析-GIS底图/CHGIS/CHCIS1911prov_gbk_shp_pts/v6_1911_prov_pts_gbk.shp”
# 省级地理点数据

chgis1911_pref_pgn <-
 “/Users/jc/Documents/6-分析-GIS底图/CHGIS/CHCIS1911prefecture_gbk_shp_pgn/v6_1911_pref_pgn_gbk.shp”
# 府级边界数据

chgis1911_county_pgn <-
 “/Users/jc/Documents/6-分析-GIS底图/CHGIS/CHCIS1911county_gbk_shp_pgn/v6_1911_cnty_pgn_gbk.shp”
# 县级边界数据
```

8.1.6 转换数据

- 导入完成后，可以在区域三`Environment`下拉的`Values`中查看到导入的数据
- 但这个时候，导入的数据并非是一个数据框，无法参与运算，因此，需要将对应数据转换为数据框
- 将导入数据转换为数据集，指定GBK编码格式，否则可能会出现乱码。如文件本身编码格式为GBK，则需要指定为GBK格式；如本身是utf-8，则需要指定为utf-8格式
- 数据导入和转换之后，即能在区域三`Environment`下拉的`Data`中查看到对应的数据集，紧接着利用导入的数据和缙绅录数据库构建记录的地理坐标

```
# 转换为地理数据

chgis1911_county_pts <- st_read(chgis1911_county_pts, options = "ENCODING=GBK")

chgis1911_prov_pgn <- st_read(chgis1911_prov_pgn, options = "ENCODING=GBK")

chgis1911_prov_pts <- st_read(chgis1911_prov_pts, options = "ENCODING=GBK")

chgis1911_pref_pgn <- st_read(chgis1911_pref_pgn, options = "ENCODING=GBK")

chgis1911_county_pgn <- st_read(chgis1911_county_pgn, options = "ENCODING=GBK")
```

02

构建地理坐标

8.2.1 准备数据

- 缙绅录数据库中“籍贯省”和“地区”两个变量提供了省、府、州、县的名称，但并不涉及地理坐标
- 查看前述生成的GIS数据集可以发现，CHGIS中1911年的地理数据提供省、府、县的地理点和地理边界以及名称
- 因此，可以利用`merge()`函数以区域名称为公共列匹配两个数据集，从而赋予缙绅录数据库地理坐标信息
- 这样，便可以得到一个不仅包含官员的个人信息，还包含官员任职地/籍贯地理信息的新数据集，进而以新数据集为基础制图

8.2.2 匹配数据

- 从缙绅录数据库中提取出1911年秋的数据与1911年地理信息数据匹配
- 由于两者年份相同，所以能够较为准确的反映1911年清代文官的地理分布状况
- 请注意，由于单独一季的数据每个人仅有一条记录，所以此处省去了个人删除重复数据的步骤
- 如果要使用多季的纪录，可参考7.2.3节的步骤删除重复记录，再匹配数据

```
# 从1900-1912公开版缙绅录数据库中提取并简化1911年秋的数据

JSL1900_1912_clean %>%
  filter(年份季节 == 1911.5) %>%
  select(阳历年份,地区,机构一,机构二,机构三,官职一,姓,名,籍贯省,籍贯县,旗分,出身一,qiren) ->
  JSL1911fall

# 匹配

JSL1911fall_2 <- merge(JSL1911fall,
                       chgis1911_county_pts,
                       by.x = "机构二", by.y = "NAME_FT")
```

匹配代码说明

- 在匹配环节运用的是CHGIS1911县级点和缙绅录数据库1911秋数据集
- 公共列采用的是任职地的县级名称
- 任职地的县级名称在缙绅录数据库中为“机构二”变量，在CHGIS1911年县级点数据集中为`NAME_FT`
- 两者数据内容和结构一致，具有匹配意义，因此作为匹配工作的公共列

匹配结果

- 查看新生成的`JSL1911fall_2`可以发现约有5000多条记录匹配成功，
- 这个数据集内包含有两个被匹配数据集的所有列，每条记录不仅有官员个人信息，还有官员任职地的地理信息，可作为制图的数据集
- 但请注意，缙绅录数据库的“机构二”变量与CHGIS1911县级点数据的数据结构一致，均为“某某县”，所以不需要处理便可以直接匹配。
- 当遇到结构不一致的数据时，比如缙绅录数据库的“籍贯县”变量，标识为”某地”，没有“县”字，这时如果直接进行匹配是不会有匹配成功的纪录，因为`merge()`函数采用的是精确匹配的方法。遇到这种情况，可以为缙绅录数据库的“籍贯县”变量增加一个“县”字便能解决

8.2.3 数据清理 – 为指定变量增减指定字符

- R可为指定变量增减指定字符
- 以“省”为例，缙绅录数据库"籍贯省"变量中省份不带“省”字，而CHGIS1911地理边界/地理点数据的文件中省份也不带“省”字，所以两者可以直接匹配
- 而缙绅录数据库中“地区”变量的省份带有“省”字，而CHGIS中没有，遇到这种情况，可以用以下代码增加或去掉“省”字，以确保匹配成功。
- 以下两条代码以“籍贯省”为例增加或去掉“省”字，请根据实际情况使用请使用者根据实际情况使用：

```
# 为指定变量增减指定字符，根据情况使用

JSL1911fall <- JSL1911fall %>% mutate(籍贯省 = ifelse(籍贯省 != "", paste0(籍贯省, "省"), 籍贯省))
table(JSL1911fall$籍贯省)

# 注意：如果不小心多次运行了以上代码，
# 可能会重复在变量上增加省字，可以用以下代码移除
JSL1911fall <- JSL1911fall %>% mutate(籍贯省 = str_remove(籍贯省, "省$"))
```

数据清理环节说明

- 数据标准化的步骤并非仅有增减字符一项，数据清理也是数据标准化的一部分
- 前述4.2节和6.2.1节分别整理了“籍贯省”和“出身一”变量
- 但缙绅录数据库的
  - “机构一”
  - “机构二”
  - “机构三”
  - “籍贯县”
- 等地理信息变量并未整理，导致数据的匹配率仅有37%
- 如果将上述变量结合数据清理方法进行精细化处理，其匹配率会得到显著提升
- 由于本教程主要介绍方法，加入冗长的整理代码会影响制图步骤的延续性，故略去了数据清理环节。使用者可根据自身需求整理上述的地理信息变量，以提高匹配率

03

绘制点密度图

8.3.1 绘制点密度图的总体步骤

- 首先绘制底图
- 再利用点密度函数将点添加到底图图层之上
- 最后设置标题、图例和调整图形的主题

8.3.2 绘制底图

- 现在需要绘制1911秋全国文官任职地分布图
- 首先需要绘制全国省级边界，可以直接利用CHGIS提供的1911年全国省级边界底图，也就是8.1节导入并转换的省级边界底图`chgis1911_prov_pgn`
- 考虑到全国省级边界比较大，可以再进一步加入1911年全国府级边界底图`chgis1911_pref_pgn`，来探讨文官在省份内部的分布状况
- 在选取地理边界底图时，尤其注意要和地理点数据区分开来。若在数据导入环节没有注意到两者的区别进而产生混淆，可以在数据集内部查看`geometry`变量判断
- 例如，府级边界数据`chgis1911_pref_pgn`数据集中`geometry`变量内容的格式为`MULTIPOLYGON(x, ...)`；
- 而县级地理点数据`chgis1911_county_pts`数据集的`geometry`变量内容的格式为`POINT(x, y)`，使用者可以结合两者进行区分

```
# 将数据集转换为sf对象以便制图

JSL1911fall_2 <- st_as_sf(JSL1911fall_2)
chgis1911_pref_pgn <- st_as_sf(chgis1911_pref_pgn)
chgis1911_prov_pgn <- st_as_sf(chgis1911_prov_pgn)
```

8.3.3 绘制点

- 前述环节给每条记录均构建了坐标，这意味着每条记录均有自己在坐标轴上的点
- 在`JSL1911fall_2`这个数据集中，每条记录点坐标的获取是依靠`geometry`变量
- 查看`JSL1911fall_2`数据集中的`geometry`变量，它的格式是`POINT(x，y)`，可以发现括号中前者代表`x`轴位置（经度），后者代表`y`轴位置（纬度）
- 这时便可以利用`sf`包中的`st_coordinates()`函数从`geometry`变量提取坐标信息
- 之所以要提取坐标而不是直接套用`geometry`变量进行制图，是因为`geometry`是一个空间点变量
- 而点密度制图函数`geom_pointdensity()`不支持直接输入空间几何对象，它需要明确的`x`和`y`数值坐标，需要通过`st_coordinates()`提取坐标后，才能传递给图层编辑函数`aes(x=..., y=...)`

```
# 绘制“1911年秋全国文官任职地分布”点密度图，浅蓝色部分为代码注释

GIS_PointDensity <- ggplot() +
  geom_sf( data = chgis1911_pref_pgn,      # 先设置府级边界作为第一层底图
         mapping = aes( geometry = geometry ), # 地理边界数据可以在图层编辑函数`aes()`直接套用`geometry`
         color = "grey60", # 使用浅灰色
         size = 1,      # 细线条
         alpha = 0.5,     # 半透明
         fill = NA ) +   # 无填充
  geom_sf( data = chgis1911_prov_pgn,      # 再设置省级边界作为第二层底图，防止府级边界覆盖省级边界
         mapping = aes( geometry = geometry ), # 地理边界数据可以在图层编辑函数`aes()`直接套用`geometry`
         color = "black",  # 深色突出
         size = 1.5,      # 粗线条
         fill = NA) +     # 无填充
  geom_pointdensity(data = JSL1911fall_2,   # 绘制点，利用8.2节构建好地理坐标的数据集
                  mapping = aes( x = st_coordinates(geometry)[,1],  # 地理点数据需要利用`st_coordinates()`函数从`geometry`
变量提取坐标信息，[,1]意为提取第一列作为经度
                                 y = st_coordinates(geometry)[,2] ), # 地理点数据需要利用`st_coordinates()`函数从`geometry`变量
提取坐标信息，[,2]意为提取第二列作为纬度
                  color = "#f47720",
                  size = 0.5, # 点大小
                  alpha = 0.6, # 点透明度
                  adjust = 0.8)  # 控制带宽，显示更精细的局部细节，该值小于1适用于观察高密度区域的微小聚集

GIS_PointDensity # 查看图片
```

8.3.4 设置主题

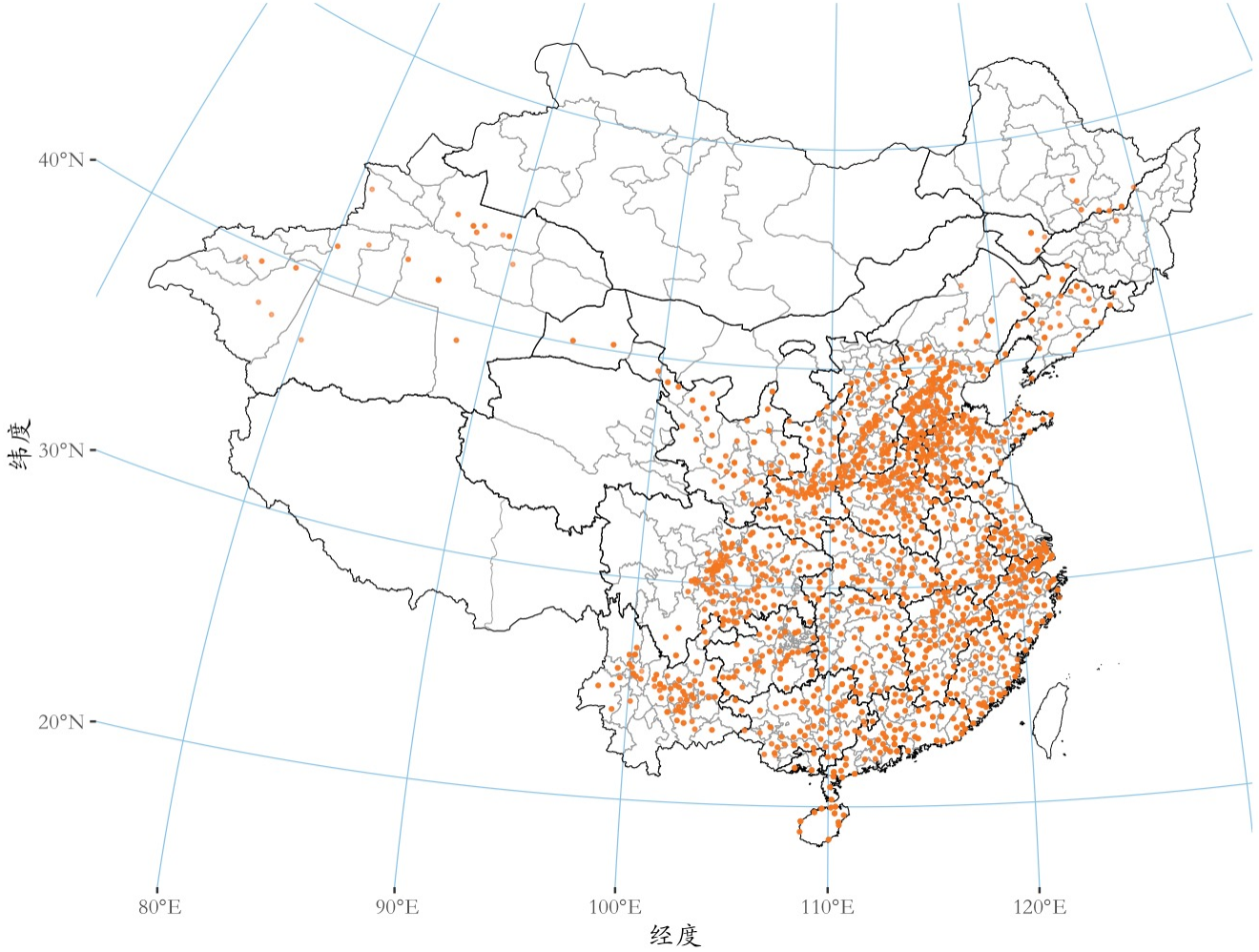
- 主要用到`labs()`设置标题和XY轴名称，`theme()`函数调整图形主题
- 函数解析在第五章

设置主题，将以下代码放到“绘制点密度图”函数后面，与其一起使用，同时利用”+”号连接

由于加入该代码会导致代码整体偏长，如未特殊说明的情况下，往后的制图函数默认使用了该代码，但不在函数中显示

```
labs(title = "1911年秋全国文官任职地分布（每个点代表一名文官）",
       x = "经度", y = "纬度") +
 theme(text = element_text(family = "STKaiti",size = 12), # 如果是MAC OS系统，请保留该代码，否则会出现字符显示不全的问题
       panel.background = element_rect(fill = "white"), # 设置面板背景颜色
       plot.background = element_rect(fill = "white"),    # 设置图片背景颜色
       plot.title = element_text(hjust = 0.5),          # 设置图片标题居中
       panel.grid.major = element_line(color = "#88c4e8", linewidth = 0.25), # 主要网格线颜色和大小
       )
```

1911年秋全国文官任职地分布（每个点代表一名文官）



图形说明

- 观察上图可以发现，1911年秋文官任职地分布较为密集的区域为华北地区和江南地区，表明清政府在这些区域设置了较多的文官
- 但这仅是一部分的匹配结果，比如青海、蒙古、西藏等地区的官员没有显示出来，这是由于`JSL1911fall`和`chgis1911_county_pts`两个数据集地名变量的不一致导致的结果
- 如果进行完全的精细化匹配，便能够得到更完整的图像以考察清代文官任职地分布状况
- 由于缙绅录数据库的数据量够大，所以点密度图在很大程度上已经能反映分布的趋势。如果遇到只需要观测数据集中趋势而不需要显示点的分布的情况，可以使用核密度图

04

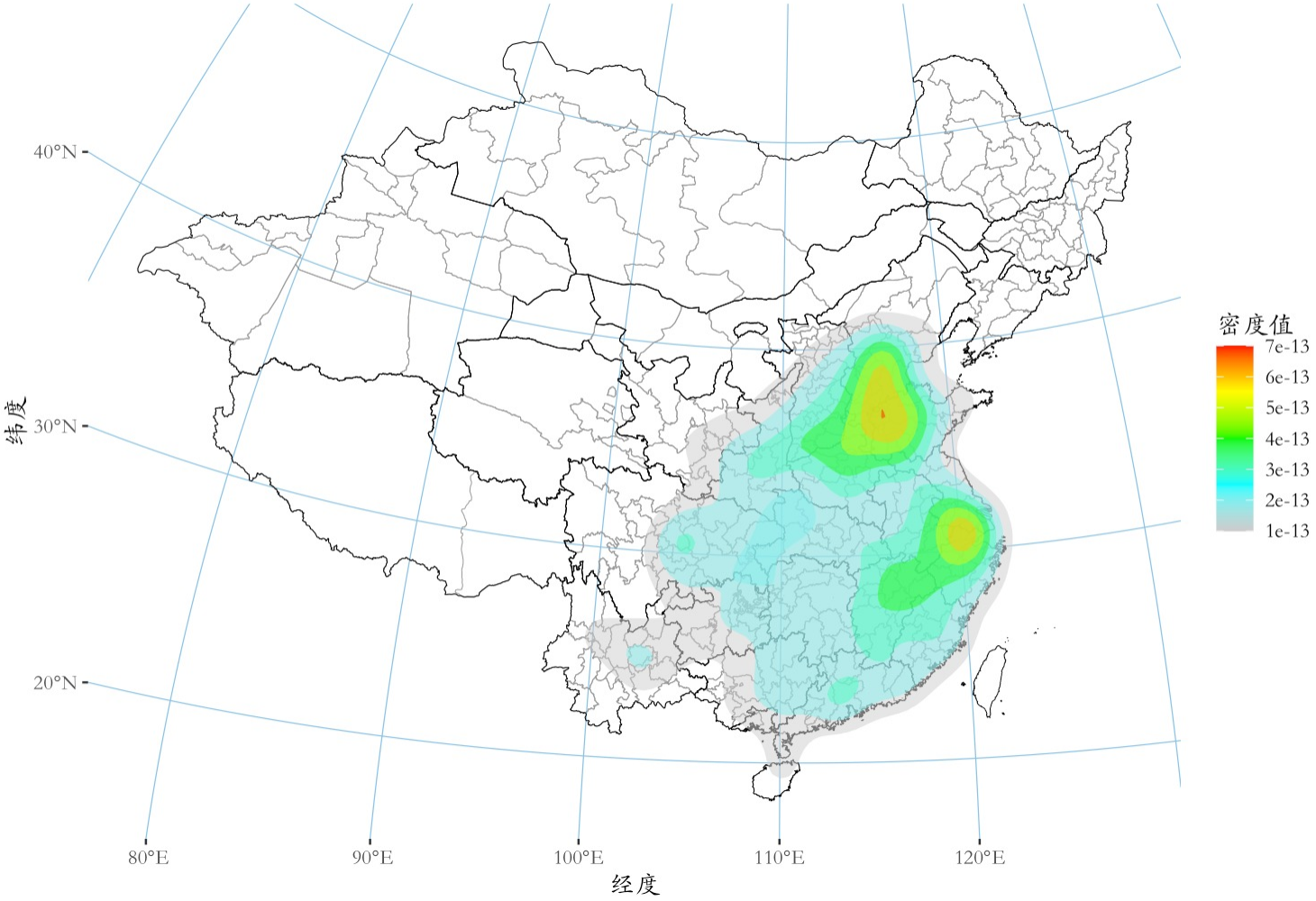
绘制核密度图

8.4.1 绘制核密度图的步骤

- 绘制核密度图的步骤与绘制点密度图的步骤基本一致
- 但需要将点密度制图函数`geom_pointdensity()`更换为核密度制图函数`stat_density_2d()`，然后调整`stat_density_2d()`函数的一些参数

```
# 绘制1911年秋全国文官任职地分布核密度图，浅蓝色部分为代码注释

GIS_KDE <- ggplot() +
  geom_sf( data = chgis1911_pref_pgn,
           mapping = aes( geometry = geometry ),
           color = "grey60",
           size = 1,
           alpha = 0.5,
           fill = NA ) +
  geom_sf( data = chgis1911_prov_pgn,
           mapping = aes( geometry = geometry ),
           color = "black",
           size = 1.5,  fill = NA) +
  stat_density_2d(
    data = JSL1911fall_2, # `stat_density_2d()`函数根据点分布自动计算核密度
    mapping = aes(x = st_coordinates(geometry)[,1],
                  y = st_coordinates(geometry)[,2],
                  fill = after_stat(level)), # 填充使用自动计算的"level"
    color = NA,
    geom = "polygon",   # 绘制填充多边形
    contour = TRUE,     # 开启等高线
    alpha = 0.5) + # 设置透明度
  scale_fill_gradientn(  # 添加颜色渐变
    colours = c("grey80", "cyan", "green", "yellow", "red"),
    name = “密度值”, trans = "identity")

GIS_KDE # 显示图片
```

1911年秋全国文官任职地分布核密度图



图像说明

- 根据以上核密度图可以发现，清政府在直隶山东一带、江浙地区设置了较多的文官
- 全国文官密度最高的区域在山东西部
- 结合黄河的走向可以发现，清末文官的分布与黄河下流沿岸地区有所重合，这或许能够说明清政府一直很重视黄河治理
- 核密度图较点密度图而言更容易观察到分布的集中趋势

05

绘制热力图

8.5.1 绘制热力图的步骤

- 热力图能够将空间划分为网格计算密度，进而通过色阶展示密度值，它的视觉效果更加直观，尤其适合高密度点数据可视化

- 热力图的绘图步骤与核密度图相似，只是要将`stat_density_2d()`函数的`geom`参数修改为`raster`，将其转换为栅格热力形式

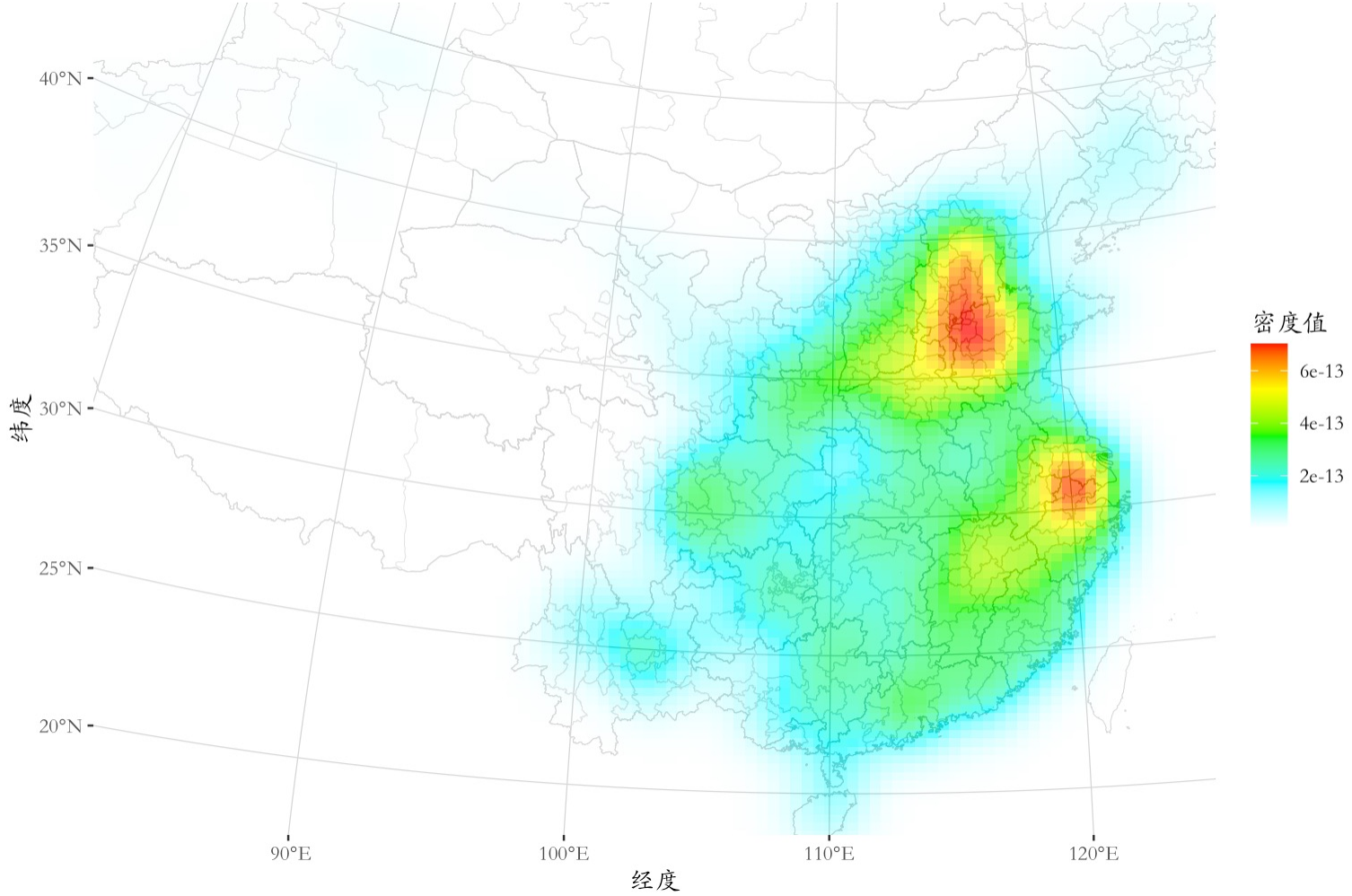
- 同时，需要将图层函数`aes()`中的`fill`参数改填充使用自动计算的`density`，即密度值而非层级

- 同时需要加入`coord_sf()`函数限定地图显示范围，`coord_sf()`是`ggplot2`专门用于地理空间数据的坐标系函数，可用于设置地图显示范围。

- `range()`函数可以计算这些 X、Y 坐标的最小值和最大值，可以设置地图的水平方向显示范围，使其刚好包含所有数据点。加入`coord_sf()`和`range()`函数可以解决热力图显示太多空白区域的问题

```
# 绘制1911年秋全国文官任职地分布热力图，浅蓝色部分为代码注释

GIS_Heatmap <- ggplot() +
  geom_sf( data = chgis1911_pref_pgn,
           mapping = aes( geometry = geometry ),
           color = “grey40”,  size = 1,  alpha = 0.5, fill = NA ) +
  geom_sf( data = chgis1911_prov_pgn,
           mapping = aes( geometry = geometry ),
           color = “black”, size = 1.5,  fill = NA) +
  stat_density_2d(
    data = JSL1911fall_2,
    mapping = aes(x = st_coordinates(geometry)[,1],
                  y = st_coordinates(geometry)[,2],
                  fill = after_stat(density)),
    geom = "raster",   # 使用栅格热力图
    contour = FALSE,    # 关闭等高线
    alpha = 0.8 ) +      # 设置透明度
coord_sf( xlim = range(st_coordinates(JSL1911fall_2)[,1]),
      ylim = range(st_coordinates(JSL1911fall_2)[,2]),
      expand = FALSE) +  # 严格使用 xlim/ylim 指定的范围，不留任何边距
  scale_fill_gradientn(  # 添加颜色渐变
    colours = c("white", "cyan", "green", "yellow", "red"), # 指定五种颜色渐变
    name = "密度值", # 图例标题
    trans = “identity”)  # 指定颜色渐变的变换，可以对颜色值进行对数变换、平方根变换等

GIS_Heatmap # 显示图片
```

1911年秋全国文官任职地分布热力图



图像说明

- 观察上图可以发现，清代文官任职地分布的高密度区域位于湖北地区和江南地区，且热力图与底图的契合度更高，能够更直观地看到高密度区域所在的府级行政单位

- 但其也容易出现分辨率不高的问题，需结合实际情况谨慎使用。

- 结合上述点密度图、核密度图和热力图的实际例子，可以归纳这三种图形在缙绅录数据库等历史数据库分析中的优势和适用场景。

    - 点密度图计算效率高，保留原始数据点，适用于较小规模数据集（在大规模数据集亦能展示分布趋势），展示微观结构
    - 热力图解读直观性强，与底图契合度高，适用于大规模数据集，展示宏观趋势
    - 核密度图处于以上二者之间，尤其适用于展示空间结构，分析单中心或多中心的分布模式。使用者可结合自身数据库的提点和具体的场景选择合适的展示形式

06

添加指北针、比例尺和图像说明

8.6.1 添加指北针、比例尺和图像说明

- 一幅完整的GIS图像除了底图和图层外，还需要添加指北针、比例尺和图像说明，帮助读者快速理解图形的方位、大小和数据来源

- 此外，还可以制作点密度图和核密度图的组合图

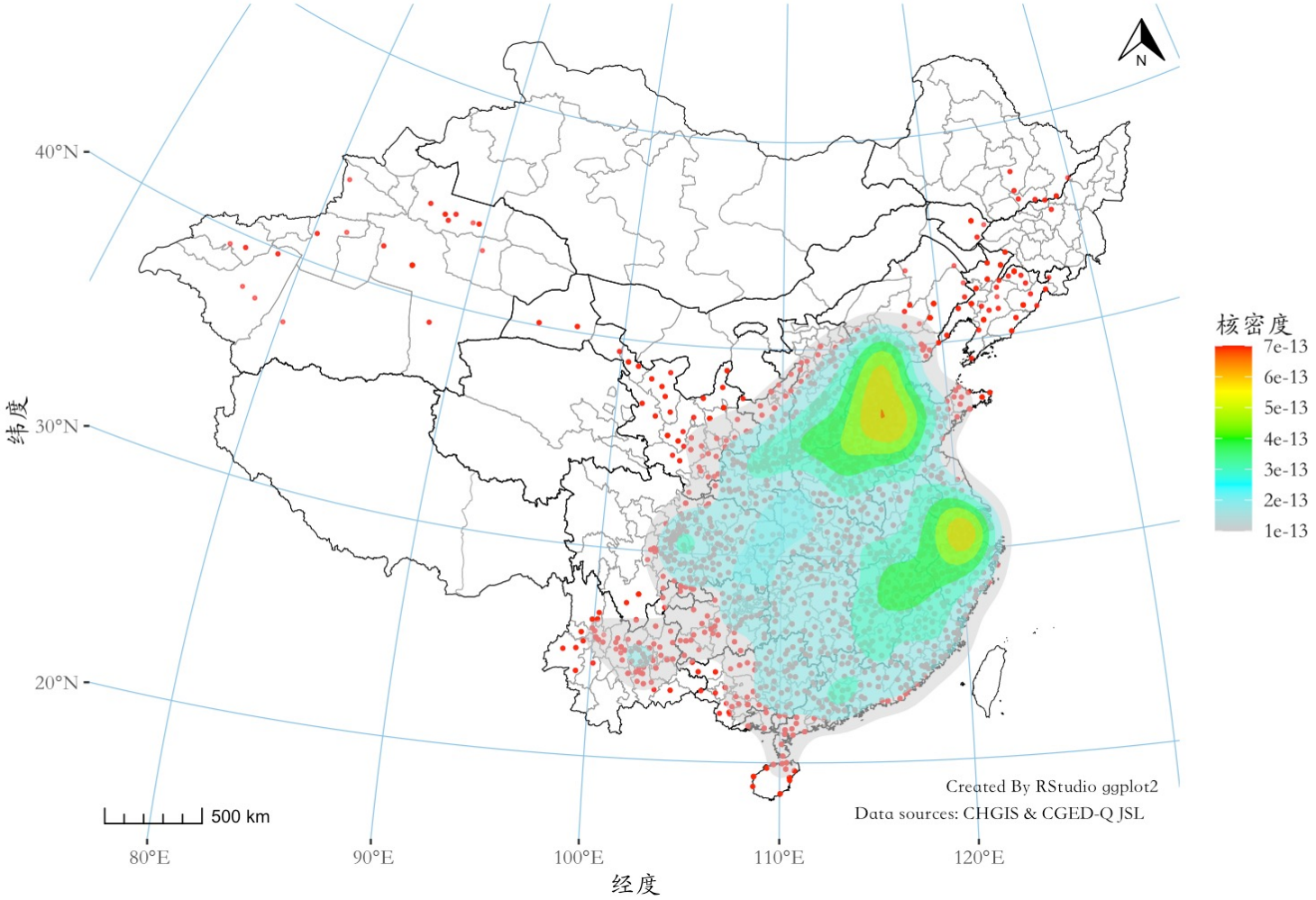
#添加指北针、比例尺和图像说明，浅蓝色部分为代码注释

由于加入该代码会导致代码整体偏长，如未特殊说明的情况下，往后的制图函数默认使用了该代码，但不在函数中显示

```
annotation_scale( # 添加比例尺
   location = "bl",  # 底部左 left
   style = "ticks",  # 刻度样式，输入"?annotation_scale"以查看其他样式
   width_hint = 0.1  # 比例尺宽度
 ) +
 annotation_north_arrow( # 添加指北针
   location = "tr",  # 顶部右 right
   style = north_arrow_orienteering (text_size = 5),
 # 指北针样式，输入"?annotation_north_arrow"以查看其他样式
   height = unit(0.75, "cm"),
   width = unit(0.75, "cm")
 ) +
 annotate( # 添加数据说明
   "text", x = Inf, y = -Inf,
   label = c("Created By RStudio ggplot2
             Data sources: CHGIS & CGED-Q JSL"),
   hjust = 1.1, vjust = -0.5,
   size = 3, color = "black", family = "Songti SC"
 )
```

```
# 制作点密度和核密度组合图，浅蓝色部分为代码注释

GIS_PointKDE <- ggplot() +
  geom_sf( data = chgis1911_pref_pgn,
         mapping = aes( geometry = geometry ),  color = “grey60”,  size = 0.5,  alpha = 1,  fill = NA ) +
  geom_sf( data = chgis1911_prov_pgn, mapping = aes( geometry = geometry ),  color = “black”,  size = 1.5,
         fill = NA) +
  geom_pointdensity(data = JSL1911fall_2,
              mapping = aes( x = st_coordinates(geometry)[,1],
                             y = st_coordinates(geometry)[,2] ),
              color = “red”, size = 0.5, alpha = 0.6, adjust = 0.8) +
  stat_density_2d(data = JSL1911fall_2,
               aes(x = st_coordinates(geometry)[, 1],
                  y = st_coordinates(geometry)[, 2],
                  fill = after_stat(level)),  # 使用计算后的层级填充
               color = NA, geom = “polygon”, contour = TRUE,  alpha = 0.5) +
  scale_fill_gradientn(
    colours = c("grey80", "cyan", "green", "yellow", "red"),
    name = "核密度",
    trans = "identity")

GIS_PointKDE # 查看图形
```

1911年秋全国文官任职地分布组合图



图像说明

- 观察上图可以发现，系统在左下角添加了比例尺，在右上角添加了指北针，在右下角添加了数据说明，同时还将核密度图层置于点密度图层之上，形成了一幅相对完整的图形。

- 完成图片的制作后，可以使用`ggsave()`函数按指定文件路径导出图形，将其运用到研究之中。

8.6.2 导出图形

```
# 导出图形

ggsave("/Users/jc/Documents/R export/image/1911年全国文官任职地分布组合图.png",
       plot = GIS_PointKDE,  #设置你自己的文件保存路径
       width = 12, height = 10,
       dpi = 600,  # 高分辨率
       bg = "white")  # 设置背景色
```

07

制作局部GIS图像

8.7 制作局部GIS图像

- 在区域性的研究中，通常用到的是局部GIS图像而非完整GIS图像

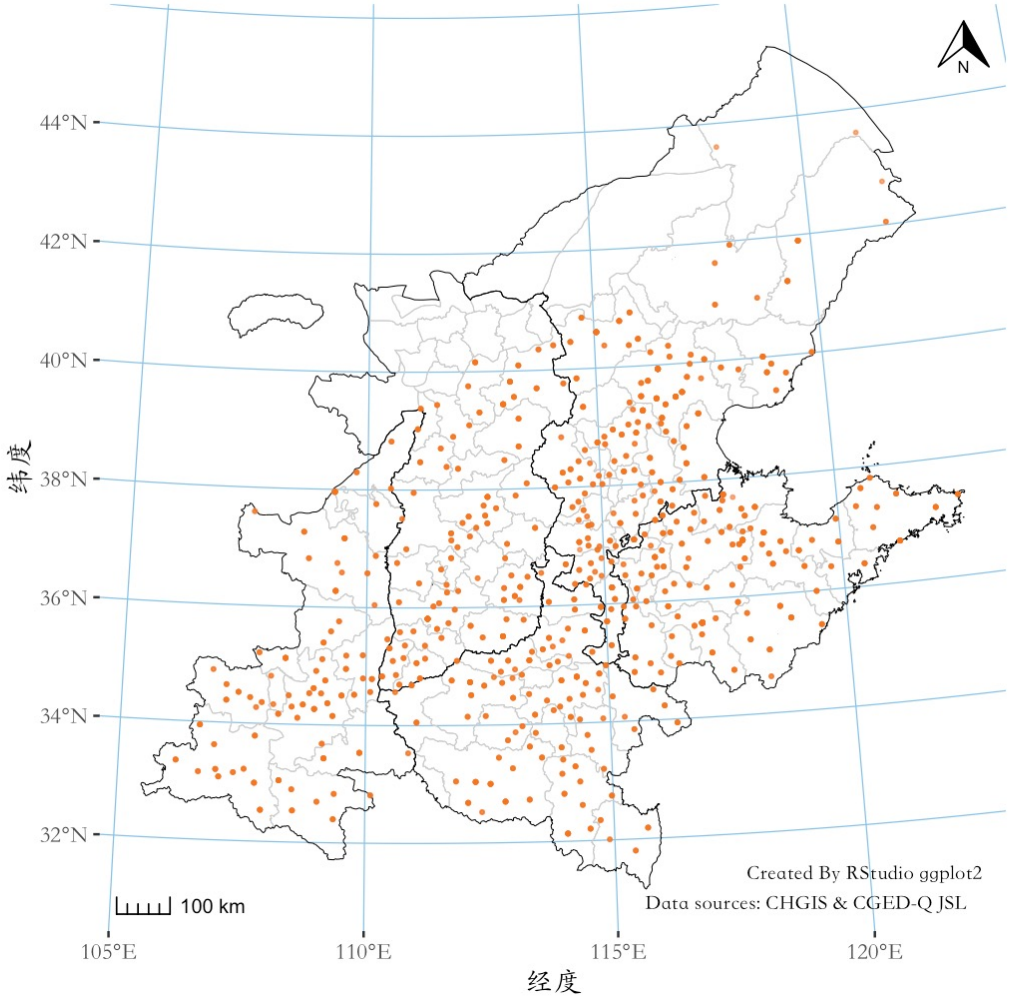
- 如何制作一个局部GIS图像？
  - 筛选出底图中的指定要素
  - 再筛选数据集中的指定要素
  - 再用筛选之后的底图和数据集来制图

- 制图步骤与全局GIS图像制图步骤一致，正确选取绘制底图和点图层的数据集后可直接复制前节使用的标题、主题、指北针、比例尺、数据说明代码。下面以华北地区（另加陕西）为例，制作清代华北地区文官任职地分布的局部GIS图像

```
# 提取shapefile中的局部文件

chgis1911_prov_pgn %>%
 filter(NAME_CH == "直隶" | NAME_CH == "山西" | NAME_CH == "山东" | NAME_CH == "河南" | NAME_CH == "陕西") ->
 chgis1911_prov_pgn_huabei
chgis1911_pref_pgn %>%
 filter(LEV1_CH == "直隶" | LEV1_CH == "山西" | LEV1_CH == "山东" | LEV1_CH == "河南" | LEV1_CH == "陕西") ->
 chgis1911_pref_pgn_huabei

# 提取件缙绅录数据库中的局部文件

JSL1911fall_2 %>%
 filter(LEV1_CH == "直隶" | LEV1_CH == "山西" | LEV1_CH == "山东" | LEV1_CH == "河南" | LEV1_CH == "陕西") ->
 JSL1911fall_2_huabei
```

```
# 制作GIS图像

GIS_Regional <- ggplot() +
  geom_sf( data = chgis1911_pref_pgn_huabei,
           mapping = aes( geometry = geometry ),
           color = "#CCCCCC",
           size = 1,
           alpha = 0.5,
           fill = NA ) +
  geom_sf( data = chgis1911_prov_pgn_huabei,
           mapping = aes( geometry = geometry ),
           color = "black",
           size = 2,
           fill = NA) +
  geom_pointdensity(data = JSL1911fall_2_huabei,
                    mapping = aes( x = st_coordinates(geometry)[,1],
                                   y = st_coordinates(geometry)[,2] ),
                    color = "#f47720",
                    size = 0.5,
                    alpha = 0.6,
                    adjust = 0.8)

GIS_Regional # 显示图片
```

1911年秋华北地区文官任职地分布点密度图



图像说明

- 上图为1911年华北地区（另加陕西）局部GIS图像，包含直隶、山东、河南、山西、陕西五省，图中一块飞地为清光绪二十九年设立的五原直隶厅，隶山西省，在今内蒙境内

- 从上图可以看出，华北地区（另加陕西）文官的分布似与省级边界有关联，陕西山西交界地区、山西河南交界地区、直隶山东河南三省交界地区官员密度较高

- 如果考虑黄河的流向，会发现华北地区（另加陕西）文官分布似与黄河走向有紧密关联，尤其在下游河道地区，官员的密度较高，这似乎表明清政府十分重视黄河的治理问题。

- 尤其需要说明的是，图中部分地区官员分布没有显示出来，如京城应分布有大量京官，这是由于在代码编写阶段为保持制图代码的简洁性和连贯性，略去了对被匹配数据集变量格式的整理环节，使用者可根据自身研究需求结合4.2节和6.2.1节整理变量的内容对局部或者全局数据作进一步精细化处理，进而提高数据的匹配率，生成更佳完整的GIS图像。

08

利用离散数据制作分级统计图

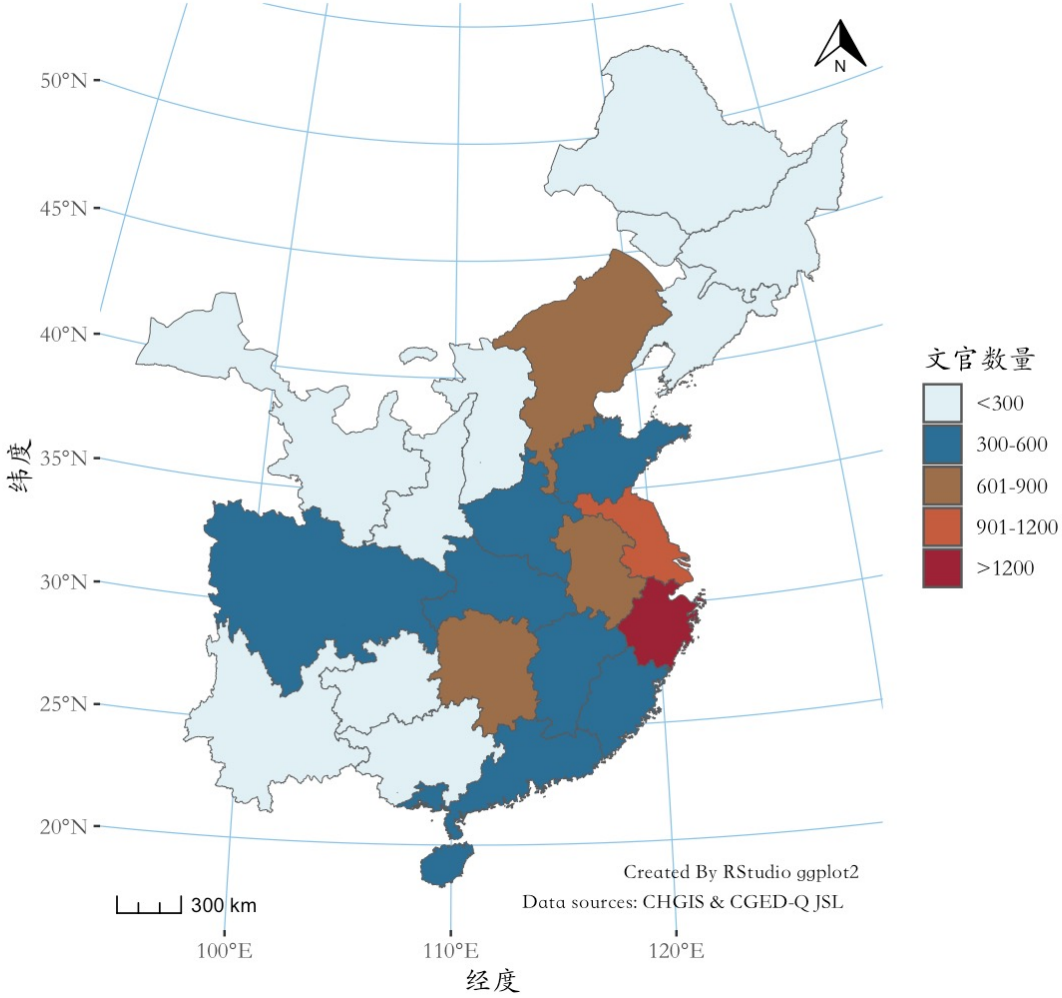
8.8.1 利用离散数据制作分级统计图

- 缙绅录数据库中没有对应的离散数据，可以转换表格生成的频数
- 比如可以利用`table()`函数统计文官省籍的分布状况
- 利用`as.data.frame()`函数将其转换为数据框
- 并将其与CHGIS省级边界数据进行匹配
- 合并为一个既带有离散数据又带有地理坐标的数据集，利用该数据集便能制作分级统计图

```
# 将表格数据转换为数据框

JSL1911fall籍贯省 <- as.data.frame(table(JSL1911fall$籍贯省))
colnames(JSL1911fall籍贯省) <- c("省份", "文官数量") # 修改列名

# 匹配shapefile文件以赋予地理边界信息

JSL1911fall籍贯省_2 <- merge(JSL1911fall籍贯省,
                    chgis1911_prov_pgn,
                    by.x = "省份", by.y = "NAME_FT")

# 制图之前生成一个新的变量对各省产生的文官数量进行分级，设置分级间隔、标签等，以便制图

JSL1911fall籍贯省_2 <- JSL1911fall籍贯省_2 %>%
  mutate(文官数量分级 = cut(文官数量, # 对数据集中的变量切分
     breaks = c(0,300,600,900,1200,1500), # 指定分级间隔
     include.lowest = TRUE,
     labels = c("<300","300-600","601-900","901-1200",">1200"))) # 设置各级别标签

# 将制图所需的文件转换为sf对象，再转换为带有坐标系统的文件

JSL1911fall籍贯省_2 <- st_as_sf(JSL1911fall籍贯省_2)
JSL1911fall籍贯省_2 <- st_transform(JSL1911fall籍贯省_2)
```

```
# 制作GIS图像完整图像（带有图例、指北针、比例尺和图像说明）

GIS_Choropleth <- ggplot(JSL1911fall籍贯省_2,aes(geometry = geometry)) +
  geom_sf( mapping = aes(fill = 文官数量分级)) + # 填充采用前述生成的分级变量
  scale_fill_manual(name = "文官数量", # 图例名称
                    values = c(  "#E0F0F5", "#2A6E96", "#9C6E4A", "#C45C3D","#9D2235" ))
```

1911年秋全国文官省籍分布图



图像说明

- 从上图中可以看出，浙江、江苏、安徽、直隶和湖南产生了较多的文官，这与清代江南地区经济、文化、教育较为发达有关

- 需要注意的是，图中未显示区域为数据不全或匹配不成功，使用者若需要更精细的匹配结果，可参考4.2节和6.2.1节整理变量的内容对“籍贯省”变量进行进一步整理，使其与CHGIS数据的匹配度更高

- 上述内容为点密度图、核密度图、热力图、组合图、局部地图和分级统计图的制作方法

- 使用者可根据研究内容和场景选择合适的可视化图形，但需要注意的是在制图之前要对缙绅录数据库的制图关键变量进行精细化处理，以满足底图数据的格式，提高匹配率和可信度。

Thanks!