

Chengdu

JACKTANSNAKE

2020/8/1

读取包

读取数据

```
Facility <- read_excel("../dataset/Facility.xlsx")
Hospital <- read_excel("../dataset/Chengdu_Hospital_All.xls")
demographics <- read_excel("../dataset/Chengdu_demographics_2019.xlsx")
```

变量名	描述
Name	机构名称
Period	创办的时间段（无用）
Year_Start	创办的年月日
City	所在城市
District	所在城市的行政区
Class	机构性质
Bed	床位数
Remaining_Bed	剩余床位数
Construction_Area	建筑面积
Floor_Space	占地面积
Address	地址
Star_Level	星级

前期准备

```
Facility_Chengdu <- Facility %>%
  filter(City == "成都") %>%
  mutate(Year_Start = mdy(Year_Start)) %>%
  mutate(Year = year(Year_Start))

Facility_Chengdu_not_imputed <- Facility_Chengdu %>%
  drop_na(Bed)

Facility_Chengdu_not_imputed %>%
  summarise(number = n())
```

```
## # A tibble: 1 x 1
##   number
##   <int>
## 1    410
```

```
Facility_Chengdu_not_imputed %>%
  summarise(total_bed = sum(Bed))
```

```
## # A tibble: 1 x 1
##   total_bed
##   <dbl>
## 1    88367
```

上面所写410家机构(包含四川省民政厅项目，按1000张床位计算)为网上有具体床位数描述的机构，共88367张有信息的床位。接下来将会对缺失床位数信息的机构根据已有建筑面积来进行推算。推算参考有床位机构的床位数与建筑面积关系。

```
Facility_Chengdu_for_ols_large <- Facility_Chengdu_not_imputed %>%
  drop_na(Construction_Area)

Facility_Chengdu_for_ols_large %>%
  summarise(number = n())
```

```
## # A tibble: 1 x 1
##   number
##   <int>
## 1    302
```

```
Facility_Chengdu_for_ols_large %>%
  summarise(total_bed = sum(Bed))
```

```
## # A tibble: 1 x 1
##   total_bed
##   <dbl>
## 1    54524
```

```
Facility_Chengdu_for_ols_large %>%
  summarise(total_Construction_Area = sum(Construction_Area))
```

```
## # A tibble: 1 x 1
##   total_Construction_Area
##   <dbl>
## 1    1918276.
```

这一步我们在筛选所有有床位数的机构中，建筑面积不为空的机构。其中共有302家机构，共54524张床位可与共1918276 m^2 的建筑面积来建立简单模型。

建立预测模型

在这一步中，我们将通过建立一个床位数与建筑面积的简单的预测模型，来为缺失床位数但有建筑面积的机构进行预测。这里，我们将建立两个模型。其一为用所有302家可用机构建立，但我们预测将会夸大每一家被预测机构的床位数（在收集数据的时候我们发现，大部分缺失床位数的机构的建筑面积都为 $1000m^2$ 左右的日间照料或是社区养老机构，本身床位数的公摊面积会大一些。而这302家里有大部分是大型的且老的护理院，养老院，其床位公摊面积都偏小。所以使用公摊面积小的数据来预测公摊面积应该比较大的机构的数据会导致床位数的夸大）。其二，为了解决第一个模型里的问题，我们考虑筛选302家机构里建筑面积比较小的来建立模型并进行预测。

大模型预测

```
set.seed(888)
Bed_Construction_Area_model_large <- train(
  Bed ~ Construction_Area,
  data = Facility_Chengdu_for_ols_large,
  method = "lm",
  trControl = trainControl(method = "cv",
                           number = 10),
  na.action = na.omit
)

summary(Bed_Construction_Area_model_large)
```

```
##
## Call:
## lm(formula = .outcome ~ ., data = dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -502.48  -61.21  -27.15   24.42  1946.62
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    7.744e+01  9.631e+00   8.041 2.08e-14 ***
## Construction_Area 1.623e-02  6.289e-04  25.808 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 152.3 on 300 degrees of freedom
## Multiple R-squared:  0.6894, Adjusted R-squared:  0.6884
## F-statistic:   666 on 1 and 300 DF,  p-value: < 2.2e-16
```

上为使用全302家机构建立的单变量模型。从统计角度来解释，建筑面积每增加 $1m^2$ ，该机构床位数将增加0.01623张。值得一提的是，在该模型中，起始床位数（当建筑面积为0时）为77张，的确符合我们会夸大床位数的判断。

```
Facility_Chengdu_for_prediction <- Facility_Chengdu %>%
  filter(is.na(Bed)) %>%
  drop_na(Construction_Area)
```

```
Facility_Chengdu_for_prediction_large <- Facility_Chengdu_for_prediction %>%
  mutate(Bed = trunc(predict(Bed_Construction_Area_model_large, newdata = Facility_Chengdu_for_prediction))) %>%
  select(Name, Bed, Year_Start, District, Class, Address, Year)
```

上两步我们在使用刚得到的模型进行预测。

```
Facility_Chengdu_imputed_large <- Facility_Chengdu %>%
  left_join(Facility_Chengdu_for_prediction_large, by = c("Name", "Year_Start", "District",
"Class", "Address", "Year")) %>%
  mutate(Bed = ifelse(is.na(Bed.x), Bed.y, Bed.x)) %>%
  select(-Bed.x, -Bed.y) %>%
  mutate(Bed = ifelse(is.na(Bed), 0, Bed)) %>%
  mutate(Address = ifelse(is.na(Address), "无", Address))

Facility_Chengdu_imputed_large %>%
  summarise(number = n())
```

```
## # A tibble: 1 x 1
##   number
##   <int>
## 1     595
```

```
Facility_Chengdu_imputed_large %>%
  summarise(total_bed = sum(Bed))
```

```
## # A tibble: 1 x 1
##   total_bed
##   <dbl>
## 1  109464
```

这一步我们在进行302家已有数据机构与被预测机构的整合。若到这一步还是没有床位数（本身没有床位数也没有建筑面积），就算为0。整合完共595家机构（应是593家，但其中有几家重名，导致整合的时候算法认为他们是不同的多出来两家），预计全成都共有109464张床位。

小模型预测

下面我们进行小模型的建立与预测。我们暂定筛选机构的标准为建筑面积小于 $3000m^2$ 。

```
Facility_Chengdu_for_ols_small <- Facility_Chengdu_not_imputed %>%
  drop_na(Construction_Area) %>%
  filter(Construction_Area < 3000)

Facility_Chengdu_for_ols_small %>%
  summarise(number = n())
```

```
## # A tibble: 1 x 1
##   number
##   <int>
## 1    144
```

```
Facility_Chengdu_for_ols_small %>%
  summarise(total_bed = sum(Bed))
```

```
## # A tibble: 1 x 1
##   total_bed
##       <dbl>
## 1       8473
```

302家机构筛选完共144家机构符合要求，共8473张床位。

```
set.seed(888)
Bed_Construction_Area_model_small <- train(
  Bed ~ Construction_Area,
  data = Facility_Chengdu_for_ols_small,
  method = "lm",
  trControl = trainControl(method = "cv",
                           number = 10),
  na.action = na.omit
)

summary(Bed_Construction_Area_model_small)
```

```
##
## Call:
## lm(formula = .outcome ~ ., data = dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -78.726 -14.520  -5.786   11.893  161.944
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      8.22161     5.69635   1.443   0.151
## Construction_Area  0.03492     0.00345  10.122 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32.73 on 142 degrees of freedom
## Multiple R-squared:  0.4191, Adjusted R-squared:  0.415
## F-statistic: 102.4 on 1 and 142 DF,  p-value: < 2.2e-16
```

上为使用144家机构建立的单变量模型。从统计角度来解释，建筑面积每增加 $1m^2$ ，该机构床位数将增加0.035张。值得一提的是，在该模型中，起始床位数（当建筑面积为0时）为8张，相较于前一个模型可能更加符合实际情况。

```
Facility_Chengdu_for_prediction_small <- Facility_Chengdu_for_prediction %>%
  mutate(Bed = trunc(predict(Bed_Construction_Area_model_small, newdata = Facility_Chengdu_for_prediction))) %>%
  select(Name, Bed, Year_Start, District, Class, Address, Year)
```

```

Facility_Chengdu_imputed_small <- Facility_Chengdu %>%
  left_join(Facility_Chengdu_for_prediction_small, by = c("Name", "Year_Start", "District",
"Class", "Address", "Year")) %>%
  mutate(Bed = ifelse(is.na(Bed.x), Bed.y, Bed.x)) %>%
  select(-Bed.x, -Bed.y) %>%
  mutate(Bed = ifelse(is.na(Bed), 0, Bed)) %>%
  mutate(Address = ifelse(is.na(Address), "无", Address))

Facility_Chengdu_imputed_small %>%
  summarise(number = n())

```

```

## # A tibble: 1 x 1
##   number
##   <int>
## 1     595

```

```

Facility_Chengdu_imputed_small %>%
  summarise(total_bed = sum(Bed))

```

```

## # A tibble: 1 x 1
##   total_bed
##   <dbl>
## 1    106130

```

同上一个模型，这里我们使用小模型来预测。同样，整合完以后成都共**595**家（应是**593**家，理由同前一个模型），共**106130**张床位，相较于前一个模型的预测少了约**3000**张床位，总体影响不大，下面所有部分都将使用小模型预测结果。

成都养老及医疗地图

```
register_google(key = "AIzaSyBvmJ3i5S_-GuGKeho8YD7ZzBG7lDXI3Go", write = TRUE)
```

```
## Replacing old key (AIzaSyBvmJ3i5S_) with new key in C:\Users\ztan1\Documents\.Renvi
```

```

set_config(
  use_proxy(url="127.0.0.1", port=1080, username="JACKTANSNAKE", password="12q12q12q")
)
df <- geocode(Facility_Chengdu_imputed_small$Address)
df_hospital <- geocode(Hospital$Address)

```

```

Facility_Complete <- cbind(Facility_Chengdu_imputed_small, df)
Hospital_Complete <- cbind(Hospital, df_hospital)

```

```

Facility_Complete <- Facility_Complete %>%
  filter(Bed < 5000)

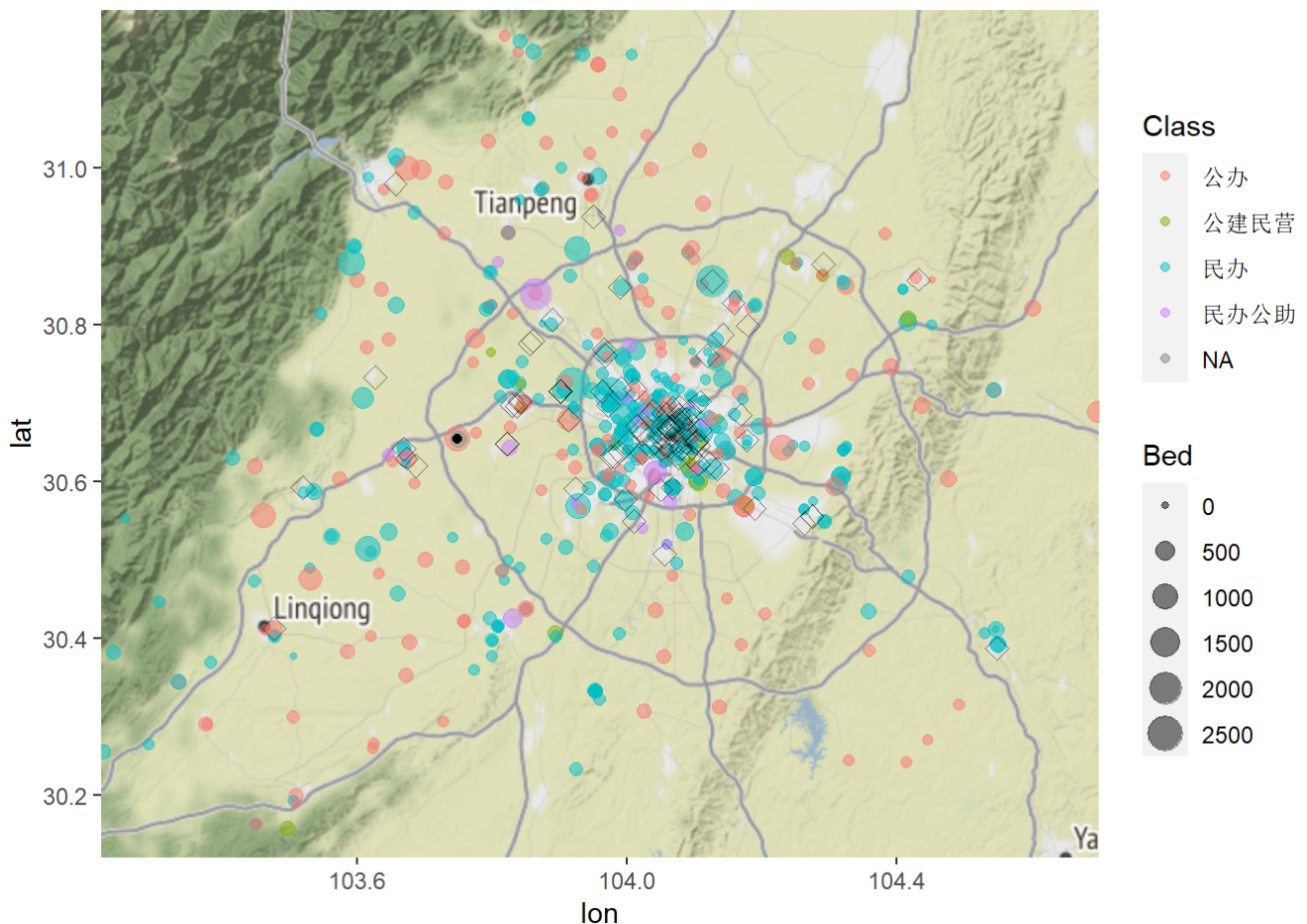
```

这里去除一个**12500**张床位老年公寓一个**5000**张床位老年公寓，因为数量过大影响后面画图。

```
bbox_Chengdu <- c(left = 103.22, bottom = 30.12, right = 104.7, top = 31.2)
map_Chengdu_satellite <- get_map(bbox_Chengdu, maptype = "satellite", source = "osm")
```

```
ggmap(map_Chengdu_satellite) +
  geom_point(data = Facility_Complete,
    aes(x = lon, y = lat, color = Class, size = Bed),
    alpha = .5, shape = 19)+
  geom_point(data = Facility_Complete,
    aes(x = 103.7494, y = 30.65515)) +
  geom_point(data = Hospital_Complete,
    aes(x = lon, y = lat), alpha = 0.5, shape = 23, size = 3)
```

```
## Warning: Removed 21 rows containing missing values (geom_point).
```



上图为成都养老及医疗机构地图。途中圆圈代表养老机构，圆圈大小代表机构床位数。菱形代表医院。中间偏左黑点代表四川省民政厅项目所在地。

图表分析

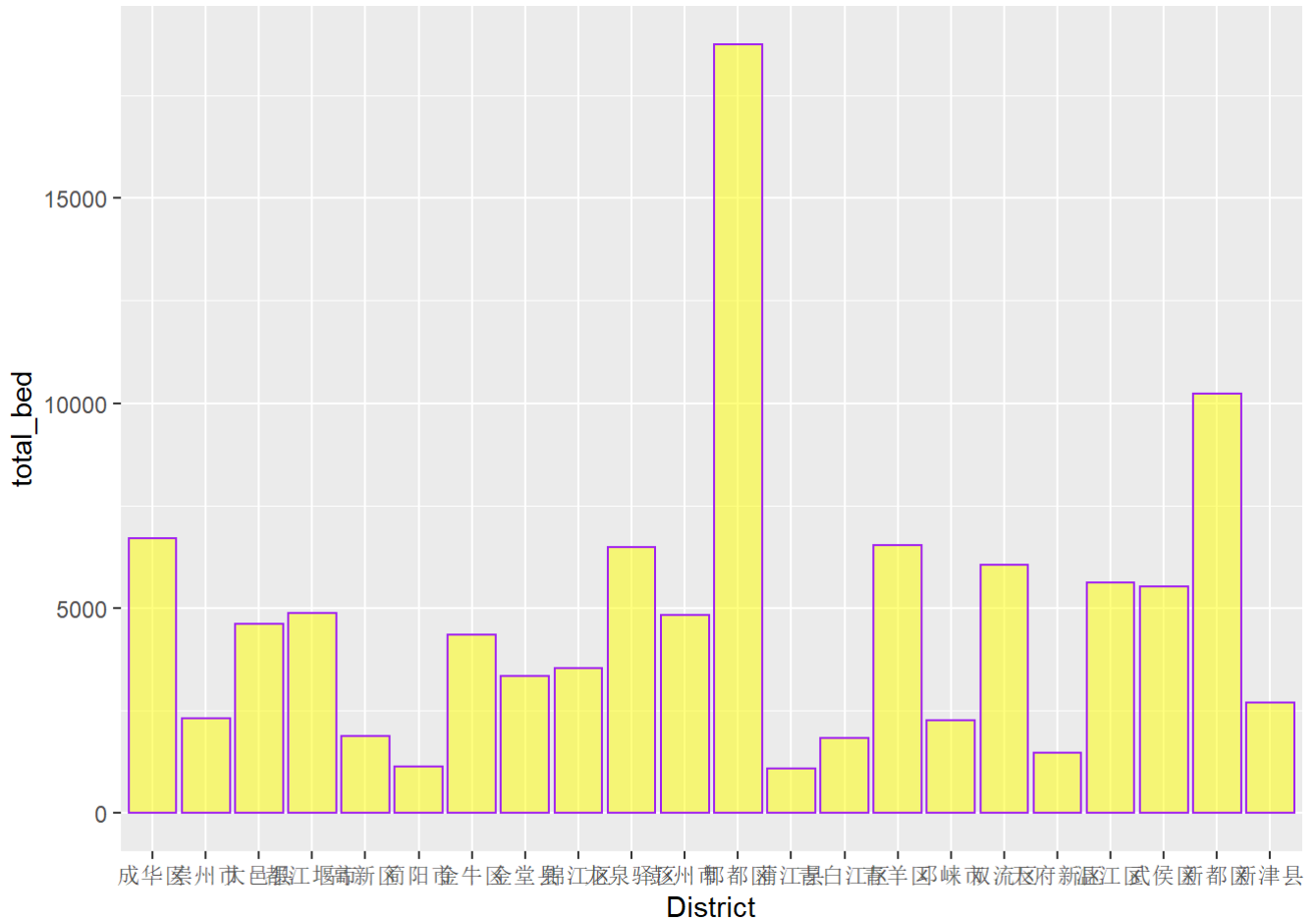
成都养老床位区块化分析

```

Facility_Chengdu_By_District <- Facility_Chengdu_imputed_small %>%
  group_by(District) %>%
  summarise(total_bed = sum(Bed))

ggplot(data = Facility_Chengdu_By_District,
  aes(x = District, y = total_bed)) +
  geom_histogram(alpha = 0.5, color = "purple", fill = "yellow", stat = "identity")

```



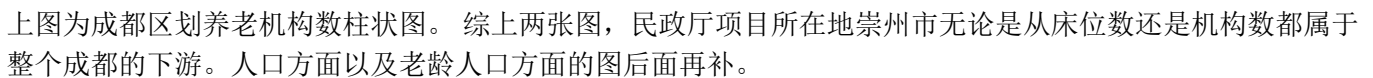
上图为成都区划养老床位柱状图。

```

Facility_Chengdu_By_District <- Facility_Chengdu_imputed_small %>%
  group_by(District) %>%
  summarise(number_facility = n())

ggplot(data = Facility_Chengdu_By_District,
  aes(x = District, y = number_facility)) +
  geom_histogram(alpha = 0.5, color = "purple", fill = "yellow", stat = "identity")

```

成都医疗机构区块化分析

```
## # A tibble: 1 x 17
##   Name Registration_Nu~ Name_Alternative Address Telephone Juridical_Name
##   <chr> <chr>           <chr>           <chr>   <chr>       <chr>
## 1 四川赫尔~ 394478786510107~ <NA>           四川省成都市~ 028-8580~ 郑元赞
## # ... with 11 more variables: PIC_Name <chr>, Date_Valid <dbl>, District <chr>,
## #   Ownership <chr>, Service_Type <chr>, Facility_Type <chr>,
## #   Facility_Level <chr>, Property <chr>, Bed <dbl>, Dentist_Chair <dbl>,
## #   Medical_catalog <chr>
```

```
Hospital %>%
  filter(Name_Alternative == "成都市老年病医院（成都市第二干部疗养院）")
```

```
## # A tibble: 3 x 17
##   Name Registration_Nu~ Name_Alternative Address Telephone Juridical_Name
##   <chr> <chr>           <chr>           <chr>   <chr>       <chr>
## 1 成都市第~ 450751864510123~ 成都市老年病医院（成都市第二干~ 成都市温江区~ 028-8272~ 雷建国
## 2 成都市第~ 450751864510123~ 成都市老年病医院（成都市第二干~ 成都市温江区~ 028-8272~ 雷建国
## 3 成都市第~ 450751864510123~ 成都市老年病医院（成都市第二干~ 成都市温江区~ 028-8272~ 雷建国
## # ... with 11 more variables: PIC_Name <chr>, Date_Valid <dbl>, District <chr>,
## #   Ownership <chr>, Service_Type <chr>, Facility_Type <chr>,
## #   Facility_Level <chr>, Property <chr>, Bed <dbl>, Dentist_Chair <dbl>,
## #   Medical_catalog <chr>
```

```
Facility_Chengdu_imputed_small %>%
  filter(Name == "四川赫尔森康复医院养老中心")
```

```
## # A tibble: 1 x 13
##   Name Year_Start City District Class Remaining_Bed Floor_Space
##   <chr> <date>      <chr> <chr>   <chr>       <dbl>       <dbl>
## 1 四川赫尔~ 2017-11-13 成都 武侯区 民办           0         31400
## # ... with 6 more variables: Construction_Area <dbl>, Address <chr>,
## #   Star_Level <lgl>, 养老地图有无 <chr>, Year <dbl>, Bed <dbl>
```

```
Hospital_Chengdu_By_District <- Hospital %>%
  group_by(District) %>%
  summarise(number_facility = n())

gerontology <- as.data.frame(grepl("老年病专业", Hospital$Medical_catalog))

rehabilitation <- as.data.frame(grepl("康复医学科", Hospital$Medical_catalog))

table(gerontology)["TRUE"]
```

```
## TRUE
##    52
```

```
table(rehabilitation)["TRUE"]
```

```
## TRUE
##    44
```

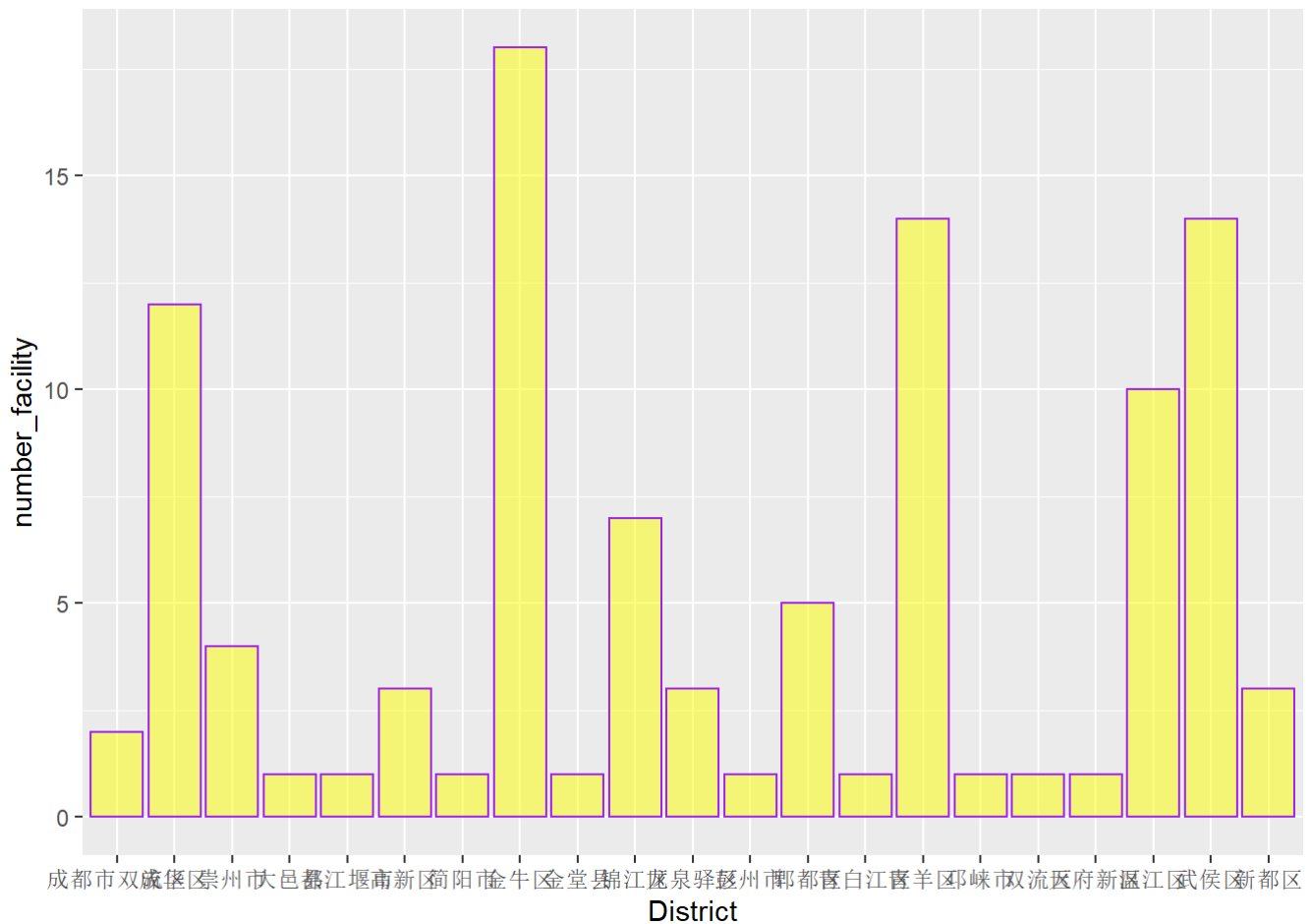
```
cbind(Hospital, gerontology) %>%
  filter(`grepl("老年病专业", Hospital$Medical_catalog)` == TRUE) %>%
  select(Name)
```

##	Name
## 1	四川省革命伤残军人医院
## 2	成都平安医院
## 3	西藏自治区人民政府驻成都办事处医院
## 4	四川现代医院
## 5	四川赫尔森康复医院
## 6	三六三医院
## 7	成都市第七人民医院1
## 8	成都市第七人民医院2
## 9	成都双楠医院
## 10	成都市第五人民医院1
## 11	成都市第五人民医院2
## 12	成都市第五人民医院3
## 13	成飞医院
## 14	成都市第三人民医院1
## 15	成都市第三人民医院2
## 16	成都市第三人民医院3
## 17	成都市第三人民医院4
## 18	成都市第三人民医院5
## 19	成都市妇女儿童中心医院1
## 20	成都市妇女儿童中心医院2
## 21	成都市妇女儿童中心医院3
## 22	成都市妇女儿童中心医院4
## 23	成都市德康医院
## 24	四川省交通运输厅公路局医院1
## 25	四川省交通运输厅公路局医院2
## 26	成都航天医院
## 27	四川友谊医院
## 28	成都市第二人民医院
## 29	成都市公共卫生临床医疗中心1
## 30	成都市公共卫生临床医疗中心2
## 31	成都市公共卫生临床医疗中心3
## 32	中国五冶集团有限公司医院1
## 33	中国五冶集团有限公司医院2
## 34	中国五冶集团有限公司医院3
## 35	成都市第八人民医院
## 36	成都中医大银海眼科医院
## 37	成都市西区医院
## 38	中铁二局集团中心医院
## 39	成都大学附属医院1
## 40	成都大学附属医院2
## 41	四川省建筑医院
## 42	成都市第四人民医院
## 43	成都市第十一人民医院1
## 44	成都市第十一人民医院2
## 45	成都市中西医结合医院1
## 46	成都市中西医结合医院2
## 47	成都新华医院
## 48	成都市第六人民医院1
## 49	成都市第六人民医院2
## 50	攀钢集团成都医院
## 51	四川省司法警官总医院
## 52	成都台双医院

```
cbind(Hospital, rehabilitation) %>%
  filter(`grepl("康复医学科", Hospital$Medical_catalog)` == TRUE) %>%
  select(Name)
```

```
##                               Name
## 1      四川省革命伤残军人医院
## 2      成都平安医院
## 3  西藏自治区人民政府驻成都办事处医院
## 4      四川现代医院
## 5      四川赫尔森康复医院
## 6      三六三医院
## 7      成都市第七人民医院1
## 8      成都市第七人民医院2
## 9      成都双楠医院
## 10     成都市第五人民医院1
## 11     成都市第五人民医院2
## 12     成都市第五人民医院3
## 13     成飞医院
## 14     成都市第三人民医院1
## 15     成都市第三人民医院2
## 16     成都市第三人民医院3
## 17     成都市第三人民医院4
## 18     成都市第三人民医院5
## 19     成都市妇女儿童中心医院1
## 20     成都市妇女儿童中心医院2
## 21     成都市妇女儿童中心医院3
## 22     成都市妇女儿童中心医院4
## 23     成都市德康医院
## 24     成都航天医院
## 25     四川友谊医院
## 26     成都市第二人民医院
## 27     中国五冶集团有限公司医院1
## 28     中国五冶集团有限公司医院2
## 29     中国五冶集团有限公司医院3
## 30     成都市第八人民医院
## 31     成都市西区医院
## 32     中铁二局集团中心医院
## 33     成都大学附属医院1
## 34     成都大学附属医院2
## 35     成都市第十一人民医院1
## 36     成都市第十一人民医院2
## 37     成都市中西医结合医院1
## 38     成都市中西医结合医院2
## 39     成都新华医院
## 40     成都市第六人民医院1
## 41     成都市第六人民医院2
## 42     核工业四一六医院1
## 43     核工业四一六医院2
## 44     攀钢集团成都医院
```

```
ggplot(data = Hospital_Chengdu_By_District,
  aes(x = District, y = number_facility)) +
  geom_histogram(alpha = 0.5, color = "purple", fill = "yellow", stat = "identity")
```

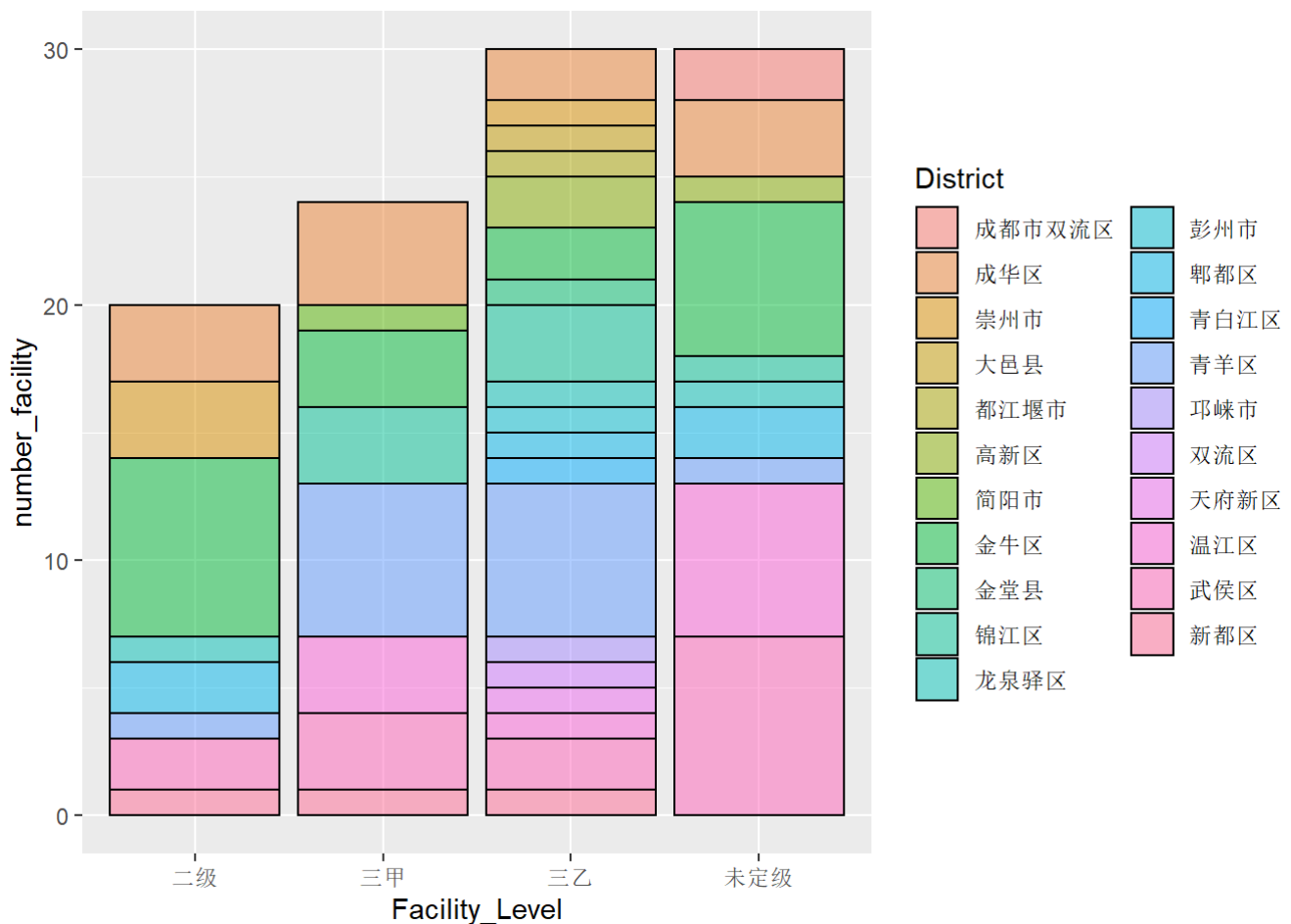


全成都市共有医疗机构104家，其中主要医疗资源集中在成都三环内的几个区（青羊区、金牛区、武侯区、成华区、温江区）。其中包括一家老年病医院位于温江区（三甲，总共有两个分院一个主院），一家康复医院位于武侯区（带有一个养老中心，15张床位）。在104家机构中，有52家医院拥有老年病专业科，以及44家医院拥有康复医学科（有大概10家三甲医院信息不在原始表上，所以不具备科室列表，大概率这几家应该都有这两个科）。带有这两个科的医院都列举在上。另医院名录可能并不是完整的，需要再跟卫计委或民政局确认一下。

```
Hospital_Chengdu_By_District <- Hospital %>%
  group_by(District, Facility_Level) %>%
  summarise(number_facility = n())

ggplot(data = Hospital_Chengdu_By_District,
  aes(x = Facility_Level, y = number_facility, fill = District)) +
  geom_histogram(alpha = 0.5, stat = "identity", position = "stack", color = "Black")
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```



从上图来看，民政厅所在崇州市医疗条件在全成都不属于特别好的，共有3家2级医院，1家三乙医院。

崇州市人口角度分析

```
Bed_Chengdu_by_District <- Facility_Chengdu_imputed_small %>%
  group_by(District) %>%
  summarise(bed = sum(Bed))

Bed_Chengdu <- Facility_Chengdu_imputed_small %>%
  summarise(bed = sum(Bed))

demographics_Complete <- demographics %>%
  left_join(Bed_Chengdu_by_District, by = "District")

population <- demographics_Complete %>%
  summarise(population = sum(Popularity))

sixty_population <- demographics_Complete %>%
  summarise(sixty_population = sum(`60_Popularity`))

cbind(Bed_Chengdu, population, sixty_population) %>%
  mutate(total_aging_percentage = sixty_population/population) %>%
  mutate(Bed_per_60 = bed/sixty_population)
```

```
##      bed population sixty_population total_aging_percentage Bed_per_60
## 1 106130 17276000 3160429 0.1829375 0.03358088
```

```
demographics_Complete %>%
  mutate(Bed_per_60 = bed/`60_Popularity`)
```

```
## # A tibble: 22 x 6
##   District `60_Popularity` Popularity aging_percentage   bed Bed_per_60
##   <chr>      <dbl>      <dbl>          <dbl> <dbl>    <dbl>
## 1 天府新区    109994    2050000      0.0537  1468    0.0133
## 2 高新区    161443    1100000      0.147   1875    0.0116
## 3 锦江区    131348     537000      0.245   3540    0.0270
## 4 青羊区    149334     841000      0.178   6538    0.0438
## 5 金牛区    196029    1211000      0.162   4359    0.0222
## 6 武侯区    144899    1083000      0.134   5528    0.0382
## 7 成华区    177988     945000      0.188   6706    0.0377
## 8 龙泉驿区   138593     859000      0.161   6498    0.0469
## 9 青白江区     94006     398000      0.236   1825    0.0194
## 10 新都区    164470     848000      0.194  10225    0.0622
## # ... with 12 more rows
```

根据网上现有人口统计数据（大概是2016年或是2017年的，年鉴上没有按区分的人口，所以人口数量上可能还少了一点），整个成都大市范围拥有总人口17276000人，其中共有316042960岁以上老年人。整个成都大市范围老龄化率在0.1829375。根据上海市出台的“9073”规则来计算60岁以上老年人口人均床位数的话，成都老人均床位数0.033，基本达标。但是细化到每个区的话，老年人人均床位数达标的基本只有成都三环以内发达的区。四川省民政厅项目所在崇州县仅有很少的人均0.014张，而该区的老龄化率也达到了全成都最高的0.24的水平。

```
ggplot(demographics_Complete, aes(x = District)) +
  geom_col(aes(y = `60_Popularity`/1000000), size = 1, color = "darkblue", fill = "white") +
  geom_line(aes(y = aging_percentage), size = 1.5, color="red", group = 1) +
  ylab("60岁以上老年人口（百万）/老龄化率") +
  xlab("区")
```

