

ccrc

JACKTANSNAKE

2020/5/20

```
library(readr)
library(tidyverse)
```

```
## -- Attaching packages -----
----- tidyverse 1.3.0 --
```

```
## √ ggplot2 3.3.0      √ dplyr   0.8.5
## √ tibble  3.0.1      √ stringr 1.4.0
## √ tidyr   1.0.2      √ forcats 0.5.0
## √ purrr   0.3.4
```

```
## -- Conflicts -----
----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggmap)
```

```
## Google's Terms of Service: https://cloud.google.com/maps-platform/terms/.
```

```
## Please cite ggmap if you use it! See citation("ggmap") for details.
```

```
library(ggthemes)
library(maps)
```

```
##
## Attaching package: 'maps'
```

```
## The following object is masked from 'package:purrr':
##
##      map
```

```
library(openintro)
```

```
## Please visit openintro.org for free statistics materials
```

```
##
## Attaching package: 'openintro'
```

```
## The following object is masked from 'package:ggplot2':
##
##   diamonds
```

```
## The following objects are masked from 'package:datasets':
##
##   cars, trees
```

```
library(tmaptools)
```

序

此数据集为加利福尼亚州的所有（？）大型老年机构的汇总，其中关键的数据包括设施的地址（已用谷歌整合为经纬坐标方便研究不同类型的机构在加州的分布情况），设施的类型（从维基百科上提供的信息来看，退休社区能分为7类，而此数据集仅分为ccrc和非ccrc两类，对于我们对于ccrc的研究有好处也有坏处），设施的容量（按unit计算）。

```
ccrc_new <- ccrc_complete %>%
  mutate(lon = lon.x) %>%
  mutate(lat = lat.x) %>%
  select(-X1, lon.x, lat.x) %>%
  drop_na(lon)
```

州数据分析

先来看一些整体的数据：

```
ccrc_new %>%
  group_by(Facility_Type) %>%
  summarise(total_facility = n())
```

```
## # A tibble: 2 x 2
##   Facility_Type          total_facility
##   <chr>                <int>
## 1 RCFE-CONTINUING CARE RETIREMENT COMMUNITY    104
## 2 RESIDENTIAL CARE ELDERLY                6925
```

从整体上来看，加州的退休型项目总共有7029家（去除70家没能找到坐标的，全部都不是ccrc），其中ccrc类型的项目有104家，非ccrc类型的有6925家。绝大部分的养老项目还是非ccrc型的。

```
ccrc_new %>%
  group_by(Facility_Type) %>%
  summarise(total_capacity = sum(Facility_Capacity))
```

```
## # A tibble: 2 x 2
##   Facility_Type          total_capacity
##   <chr>                <dbl>
## 1 RCFE-CONTINUING CARE RETIREMENT COMMUNITY    31464
## 2 RESIDENTIAL CARE ELDERLY                138653
```

具体到床位，在加州总共有170117的容量（应该是按unit来算的）。其中31464张属于ccrc类型的项目，138653属于非ccrc类型的项目。根据维基百科给出的数据，虽然美国国内的ccrc项目体量区间变化很大，但全国的平均水平为：每个机构有330个unit（这里的unit应该指的是房间），其中有231间独立或congregate（我理解为两人住），34张辅助生活床位，以及70张护理床位。也就是说，平均一个项目大致的床位产出可以粗略估计为： $\frac{2}{3} * 231 * 2 + \frac{1}{3} * 231 * 1 = 385$ 张独立床位，34张辅助床位，以及70张护理床位。百分比值为：78.7% 独立老人，7% 半自理老人，14.3% 护理老人。按照这个组成情况来看，加州的ccrc项目中共有24762张独立老人床位，2202张半自理老人床位，以及4500张纯护理床位。为了方便对比，这里引入一些加州的人口数据：加州大约有人口39512000。根据斯坦福给出的2010年的统计数据，加州65岁以上老人约占总人口的11%，而全国范围约为13%。以2020年占13%的总人口来计算，加州大约有5136560人65岁以上人口。

```
ccrc_new %>%
  group_by(Facility_Type) %>%
  summarise(avg = mean(Facility_Capacity))
```

```
## # A tibble: 2 x 2
##   Facility_Type      avg
##   <chr>          <dbl>
## 1 RCFE-CONTINUING CARE RETIREMENT COMMUNITY 303.
## 2 RESIDENTIAL CARE ELDERLY                20.0
```

分布到每个机构来看，ccrc项目平均有302个unit，低于维基百科所述的330，原因我理解可能有两点：1.加州多为海岸线，高端项目多，为了满足高端需求，刻意降低每个项目的容量，提升质量。2.加州早年便就是美国较为繁荣的州，所以养老项目大多都开发的较早，规划可能与现如今的情况有出入。在价格上暂时并未能找到加州ccrc项目的价格，但是美国全国范围的平均水平的价格如下图所示：

Paying for Peace of Mind

Here's a representative array of what you might pay at a continuing-care retirement community. Some CCRCs charge considerably more.

CONTRACT TYPE	STUDIO FOR ONE PERSON		TWO-BEDROOM UNIT FOR TWO	
	ENTRY FEE	MONTHLY CHARGE	ENTRY FEE	MONTHLY CHARGE
LIFE CARE Includes assisted living and/or nursing care as needed	\$159,700	\$2,450	\$289,450	\$4,750
MODIFIED FEE FOR SERVICE Provides some covered or discounted care	\$94,950	\$2,350	\$215,500	\$4,450
FEE FOR SERVICE No prepayment for future care	\$90,100	\$2,100	\$203,600	\$3,900

NOTE: Figures are for contracts that provide declining-balance refunds. **SOURCE:** MyLifeSite.net

浅蓝色为单人间，深蓝色为双人间双人住。数据均为左入门费，右月费。最左为三种不同的合约模式，从上到下依次为：终生保障（入门费最高，月费最高，享受生活助理或是护理（应该包含大部分医疗）），加强版付费（比全付费贵但享受小部分终身保障用户的服务），以及全付费（只含住和吃，护理医疗全部自费）。值得注意的是美国的ccrc项目的应该是不带产权的，因为在维基百科上有这样一句话：“关于ccrc值得注意的一点是，在住户选择离开社区或是离世的情况下，机构可能根据情况予以部分入门费的返还甚至不返还。”

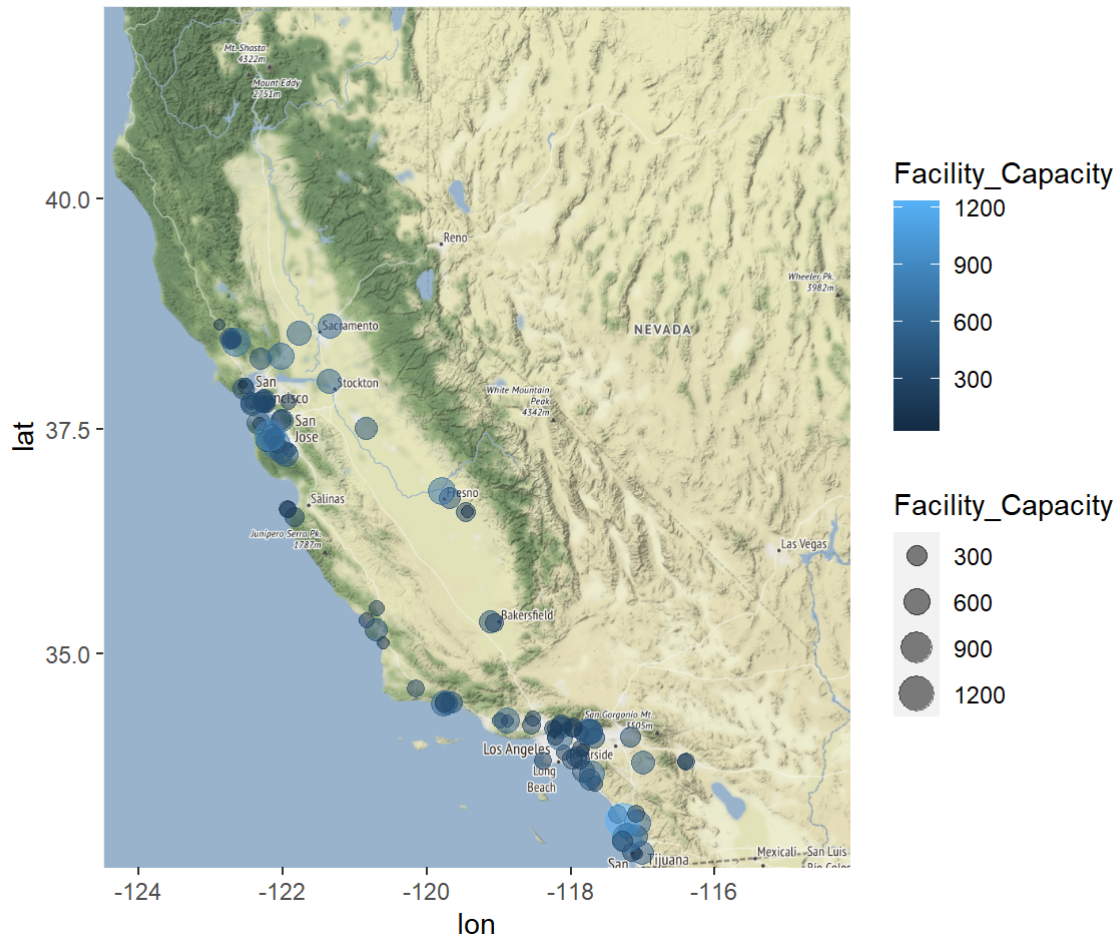
```
ccrc_new %>%  
  group_by(County_Name) %>%  
  summarise(Count = n()) %>%  
  arrange(desc(Count))
```

```
## # A tibble: 51 x 2  
##   County_Name    Count  
##   <chr>         <int>  
## 1 LOS ANGELES    1372  
## 2 ORANGE         956  
## 3 SAN DIEGO      558  
## 4 SACRAMENTO     535  
## 5 RIVERSIDE      470  
## 6 CONTRA COSTA   358  
## 7 SAN BERNARDINO 246  
## 8 SANTA CLARA    246  
## 9 ALAMEDA        241  
## 10 SAN MATEO     238  
## # ... with 41 more rows
```

此为每个郡拥有的项目的总体个数，以洛杉矶为例，将在第四部分进行城市的分析。

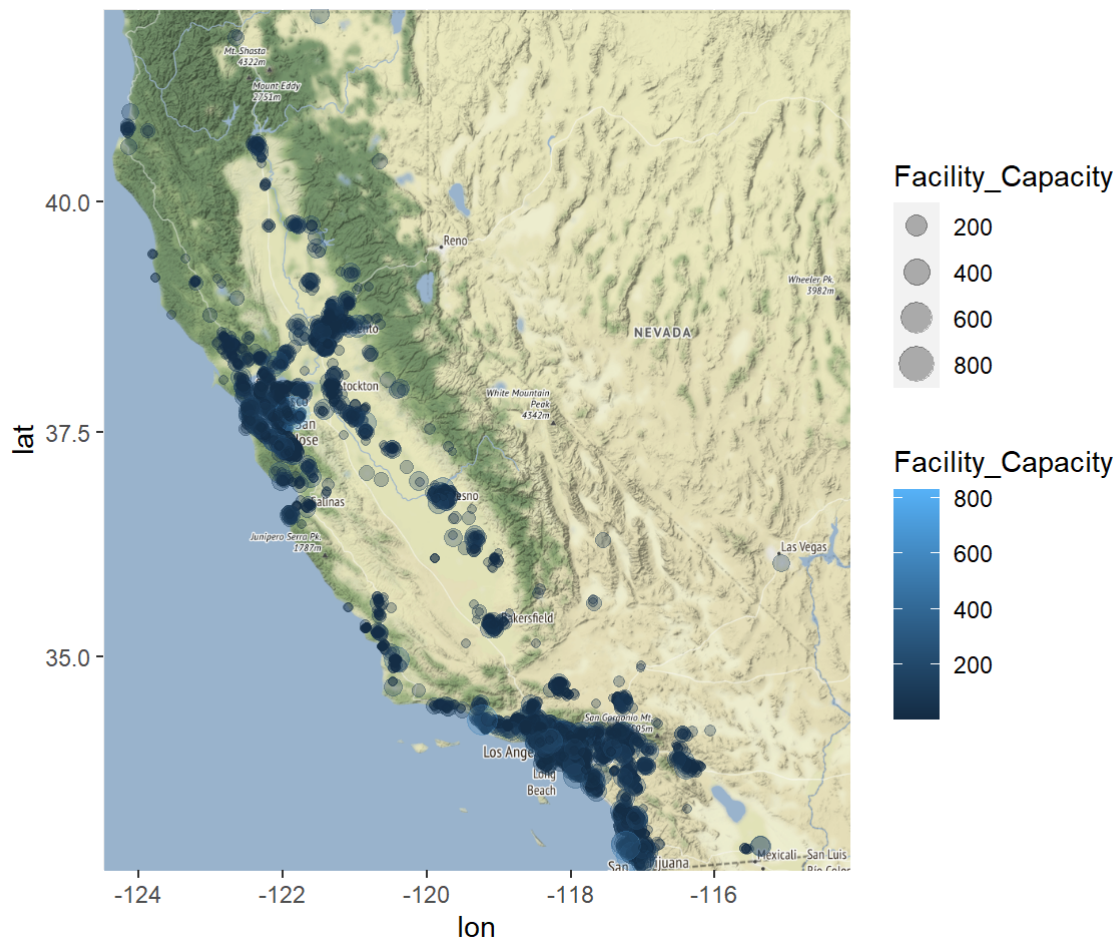
州地图分析

```
bbox_CA <- get_location_bbox("California")  
map_CA <- get_map(location = bbox_CA)  
ggmap(map_CA) +  
  geom_point(data = ccrc_new_ccrc,  
             aes(x = lon, y = lat, color = Facility_Capacity, size = Facility_Capacity),  
             alpha = .5)
```



上图为ccrc项目在加州的分布图。可以看见最大的两个1000单元左右的设施位于旧金山以及洛杉矶（算上圣地亚哥）两个加州最大的都市附近，其余项目大多都分布在加州绵长的海岸线上，偏内陆的都聚集在大的郡周围（萨克拉门托等）。

```
ggmap(map_CA) +
  geom_point(data = ccrc_new_none_ccrc,
    aes(x = lon, y = lat, color = Facility_Capacity, size = Facility_Capacity),
    alpha = .3)
```



上图为非ccrc项目在加州的分布图。从地域上大致可以分为南北两个大的聚集区，旧金山和洛杉矶（算上圣地亚哥）。比起ccrc，非ccrc类型的项目在内陆也有不小的聚集。

城市数据分析（洛杉矶）

```
ccrc_LA <- ccrc_new %>%
  filter(County_Name %in% c("LOS ANGELES"))
```

先来看一些整体的数据：

```
ccrc_LA %>%
  group_by(Facility_Type) %>%
  summarise(total_facility = n())
```

```
## # A tibble: 2 x 2
##   Facility_Type          total_facility
##   <chr>                <int>
## 1 RCFE-CONTINUING CARE RETIREMENT COMMUNITY      19
## 2 RESIDENTIAL CARE ELDERLY          1353
```

从整体上来看，洛杉矶的退休型项目总共有1374家，其中ccrc类型的项目有19家，非ccrc类型的有1353家。绝大多数的养老项目还是非ccrc型的。

```
ccrc_LA %>%
  group_by(Facility_Type) %>%
  summarise(total_capacity = sum(Facility_Capacity))
```

```
## # A tibble: 2 x 2
##   Facility_Type          total_capacity
##   <chr>                <dbl>
## 1 RCFE-CONTINUING CARE RETIREMENT COMMUNITY    4551
## 2 RESIDENTIAL CARE ELDERLY                    31253
```

具体到床位，在洛杉矶总共有35804的容量（应该是按unit来算的）。其中4551张属于ccrc类型的项目，31253属于非ccrc类型的项目。按照第二部分的计算方式来看，洛杉矶的ccrc项目中共有3582张独立老人床位，318张半自理老人床位，以及650张纯护理床位。为了方便对比，这里引入一些洛杉矶的人口数据：洛杉矶2018年大约有人口3,990,456。2018年估计的65岁以上老年人百分比为12.1%，由此统计洛杉矶大约有482845的65岁以上的老年人。

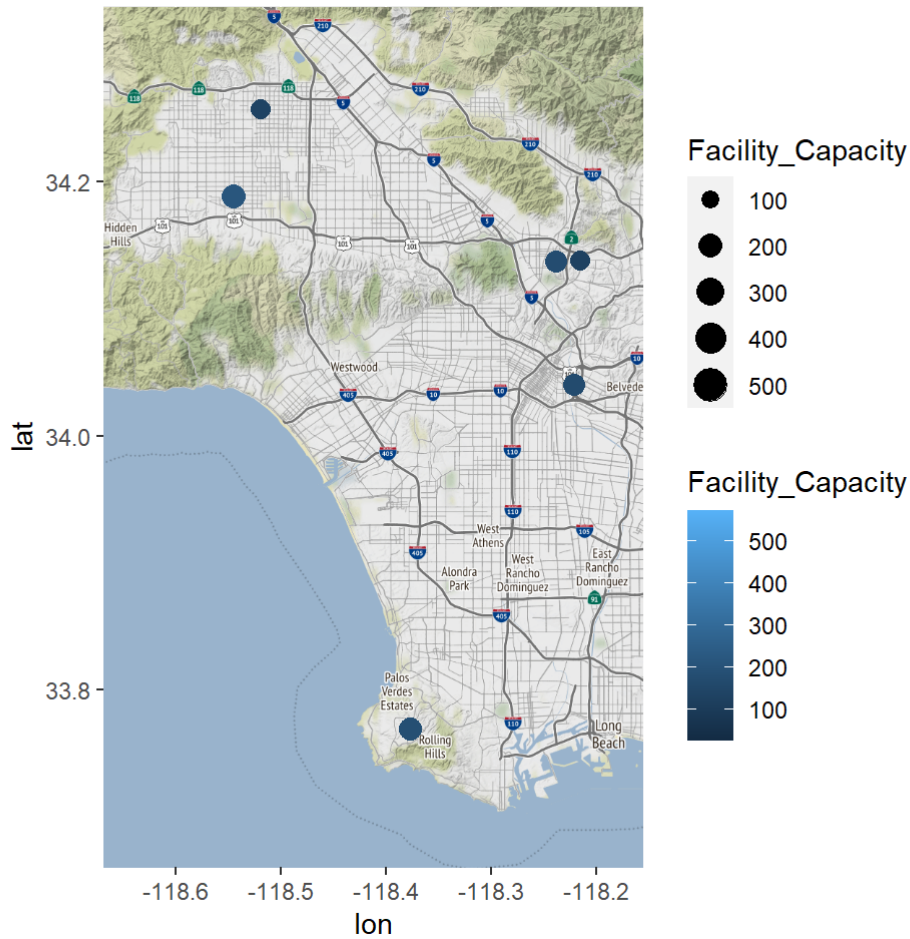
```
ccrc_LA %>%
  group_by(Facility_Type) %>%
  summarise(avg = mean(Facility_Capacity))
```

```
## # A tibble: 2 x 2
##   Facility_Type          avg
##   <chr>                <dbl>
## 1 RCFE-CONTINUING CARE RETIREMENT COMMUNITY 240.
## 2 RESIDENTIAL CARE ELDERLY                23.1
```

由上数据可见洛杉矶的平均ccrc单位数是不到加州平均的。少了将近60单元。

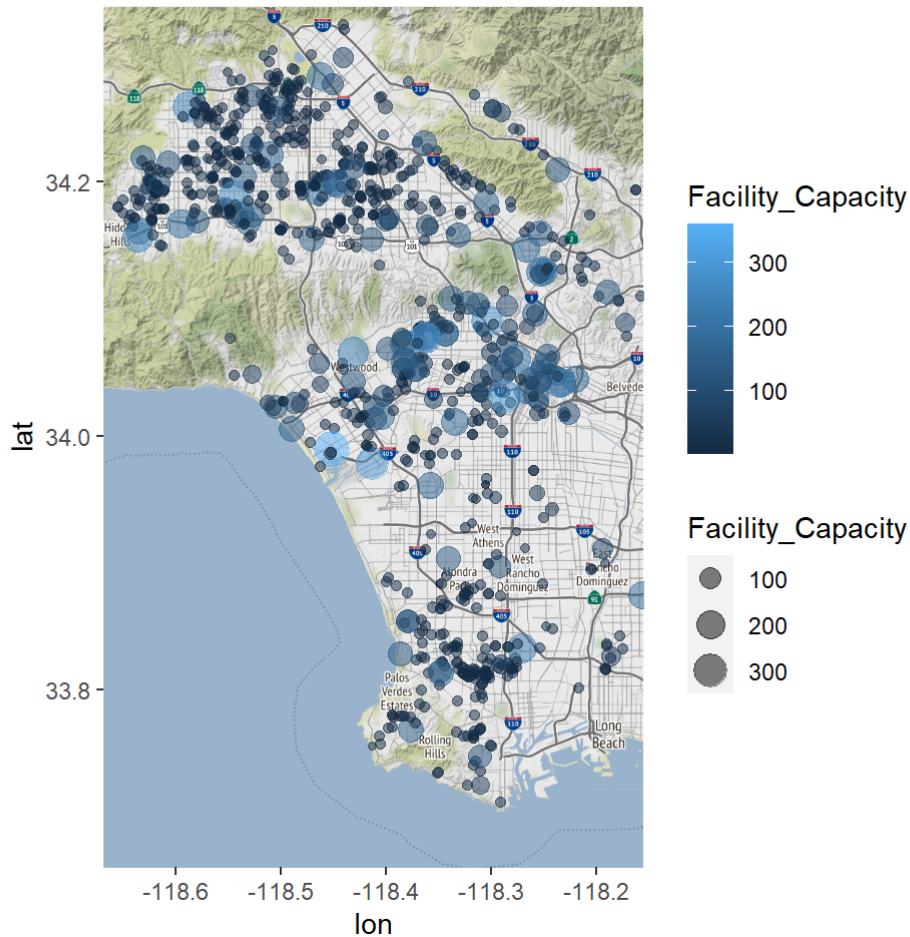
洛杉矶地图分析

```
bbox_LA <- get_location_bbox("Los Angeles")
map_LA <- get_map(location = bbox_LA)
ggmap(map_LA) +
  geom_point(data = ccrc_LA_ccrc,
             aes(x = lon, y = lat, color = Facility_Capacity, size = Facility_Capacity))
```

洛杉矶市区大约只有6家ccrc项目，剩余13家均位于城外（地图没法显示）。

```
ggmap(map_LA) +  
  geom_point(data = ccrc_LA_none_ccrc,  
             aes(x = lon, y = lat, color = Facility_Capacity, size = Facility_Capacity),  
             alpha = .5)
```

与ccrc项目相比，非ccrc型项目虽然也有430家不在室内，但是市内分布923家，城市聚集明显。

结尾

此分析只从分布出发。因为获得的数据有限，没法从地价，设施条件，非ccrc项目细分的方向来开发数据分析，因此仅供参考。

前期使用谷歌API获取经纬度代码

```
#register_google(key = "AIzaSyBvmJ3i5S_-GuGKeho8YD7ZzBG7LDXI3Go", write = TRUE)
```

```
#ccrc <- ccrc %>%
  #mutate(Facility_Type = as.factor(Facility_Type)) %>%
  #distinct(Facility_Name, .keep_all = TRUE) %>%
  #mutate(Facility_Address = paste(Facility_Address, County_Name, sep = ", ")) %>%
  #filter(Facility_Status != "CLOSED") %>%
  #select(-Closed_Date)
```

```
#library(httr)
#set_config(
  # use_proxy(url="127.0.0.1", port=1080, username="JACKTANSNAKE", password="12q12q12q")
#)
#df <- geocode(ccrc$Facility_Address)
```

```
#ccrc_full <- cbind(ccrc, df)
```

```
#ccrc_ccrc <- ccrc %>%  
  #filter(Facility_Type == "RCFE-CONTINUING CARE RETIREMENT COMMUNITY")  
#df_ccrc <- geocode(ccrc_ccrc$Facility_Address)
```

```
#ccrc_ccrc <- ccrc %>%  
  #filter(Facility_Type == "RCFE-CONTINUING CARE RETIREMENT COMMUNITY")  
#ccrc_ccrc <- cbind(ccrc_ccrc, df_ccrc)
```

```
#ccrc_complete <- full_join(ccrc_ccrc,  
  #ccrc_full, by = c("Facility_Name", "Facility_Type", "Facility_Num  
ber", "Licensee", "Facility_Administrator", "Facility_Address", "Facility_City", #"Facility_S  
tate", "Facility_Zip", "County_Name", "Regional_Office", "Facility_Capacity", "Facility_Statu  
s", "License_First_Date")) %>%  
  #mutate(lon.x = ifelse(is.na(lon.x), lon.y, lon.x)) %>%  
  #mutate(lat.x = ifelse(is.na(lat.x), lat.y, lat.x)) %>%  
  #select(-lon.y, -lat.y, -`Facility_Telephone Number.x`, -`Facility_Telephone Number.y`)
```

```
#write.csv(ccrc_complete, "ccrc_complete.csv")
```