607 Project 2 - UK Renewable Generation

Jose Mawyin

10/5/2019

UK Energy Generation Data

We will take a look at the breakdown of Renewable Generators ("Anaerobic digestion", "Hydro", "Micro CHP", "Photovoltaic", "Wind") in the different regions of the UK.

```
UK.Renewables <- as_tibble(read.csv("https://raw.githubusercontent.com/JMawyin/MSDS2019-607/master/UKRe
```

As we can see below, the dataset contains a mix of numeric and factor information as loaded into the dataframe.

```
str(UK.Renewables)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    19723 obs. of  15 variables:
##  $ FITID                     : Factor w/ 19543 levels "FIT00000001",..: 743 1867 5937 5981 7521 9587
##  $ PostCode                  : Factor w/ 2339 levels "AB","AB12","AB15",..: 4 6 5 3 2 1 3 3 21 9 ...
##  $ TechnologyTypeName        : Factor w/ 5 levels "Anaerobic digestion",..: 4 4 4 4 5 5 4 4 4 4 ...
##  $ InstallationTypeName      : Factor w/ 4 levels "Community","Domestic",..: 2 2 2 2 2 4 2 2 2 2 ...
##  $ InstalledCapacity         : num  3.96 1.5 2 2.1 50 11 1.72 3.69 4.32 2.59 ...
##  $ DeclaredNetCapacity       : num  3.96 1.5 2 2.1 50 11 1.72 3.69 4.32 2.38 ...
##  $ ApplicationDate           : Factor w/ 275 levels "10/1/10","10/10/10",..: 181 136 93 93 93 93 267
##  $ CommissionedDate          : Factor w/ 1509 levels "1/1/00","1/1/01",..: 847 58 1379 384 730 178 14
##  $ ExportStatusTypeName      : Factor w/ 5 levels "Export (deemed)",..: 1 1 1 1 2 1 1 1 1 1 ...
##  $ TariffCode                : Factor w/ 20 levels "AD/0-500/01",..: 9 9 13 9 20 16 9 9 12 9 ...
##  $ Description               : Factor w/ 20 levels "Anaerobic Digestion (<=500kW)-2010/11",..: 11 11
##  $ CountryName               : Factor w/ 4 levels "England","NULL",..: 3 3 3 3 3 3 3 3 3 3 ...
##  $ GovernmentOfficeRegionName: Factor w/ 10 levels "East Midlands",..: 6 6 6 6 6 6 6 6 6 6 ...
##  $ LocalAuthorityName        : Factor w/ 379 levels "Aberdeen City",..: 1 1 1 1 1 1 1 1 2 2 ...
##  $ AccreditationNo           : Factor w/ 19662 levels "FAD00003EN","FAD00004EN",..: 1724 897 15223 15
```

Below we can see the different variables contained in the dataset.

```
colnames(UK.Renewables)
```

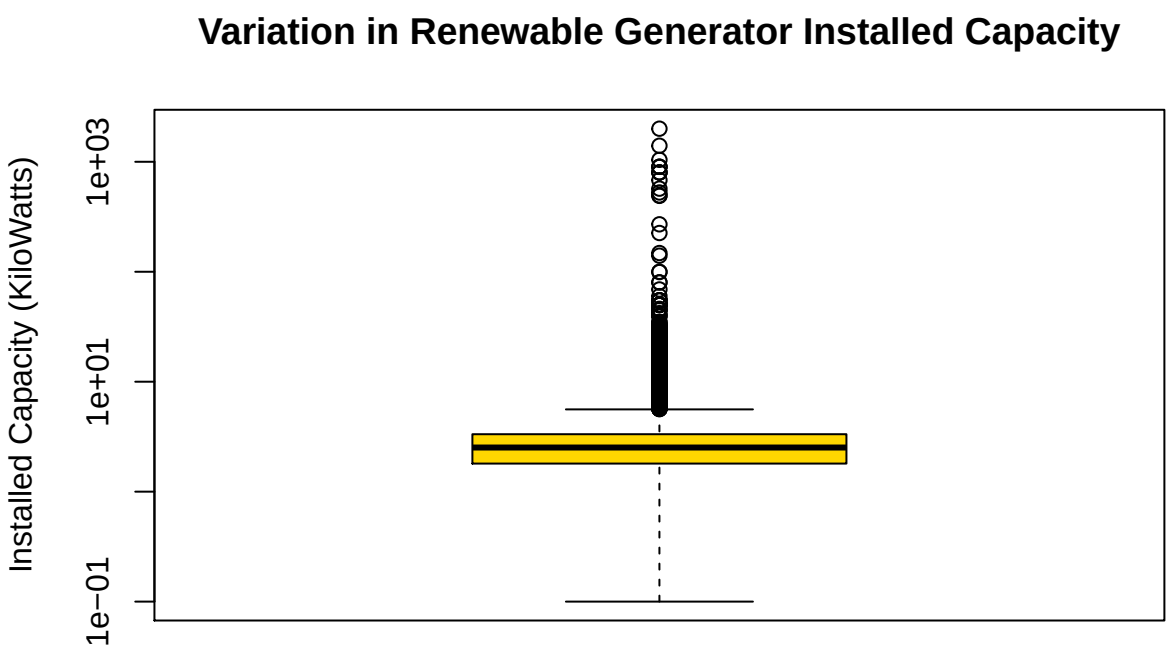
```
##  [1] "FITID"                      "PostCode"
##  [3] "TechnologyTypeName"         "InstallationTypeName"
##  [5] "InstalledCapacity"          "DeclaredNetCapacity"
##  [7] "ApplicationDate"            "CommissionedDate"
##  [9] "ExportStatusTypeName"       "TariffCode"
## [11] "Description"                "CountryName"
## [13] "GovernmentOfficeRegionName" "LocalAuthorityName"
## [15] "AccreditationNo"
```

How spread out are the installations based on their rated capacity? As we can see below, most installations are rated near 3 kilowatts. However, we have extreme outliers such as a system that was rated at 2000 kilowatts or almost 3 orders of magnitude larger than the typical installation size in kilowats.

```
summary(UK.Renewables$InstalledCapacity)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##    0.000    1.800    2.520    3.669    3.330 2000.000
```

```
Install.Cap <- UK.Renewables$InstalledCapacity
Install.Cap[(Install.Cap == 0)] <- 1
boxplot(Install.Cap, log = "y", col = "gold",  ylab="Installed Capacity (KiloWatts)",main="Variation in
```



We can see below that most of the installations are of the domestic type, fllowed by commercial and community owned PV systems.

```
table(UK.Renewables$InstallationTypeName)
```

```
##
##                 Community                  Domestic
##                       276                     19104
## Non Domestic (Commercial) Non Domestic (Industrial)
##                       321                        22
```

Most of the installations are Photovoltaic per technology and domestic per installation type.

```
T.Tech.Installs <- table(UK.Renewables$TechnologyTypeName)
```

**What is the breakdown of installations per Region and per Technology Type?

```
ByRegion.Tech <- table(UK.Renewables$GovernmentOfficeRegionName,UK.Renewables$TechnologyTypeName)
ByRegion.Tech
```

```
##
##                            Anaerobic digestion Hydro Micro CHP
##   East Midlands                              0     7         2
##   East of England                            0     6         4
##   London                                     0     0         1
##   North East                                 0     4         0
##   North West                                 0    11         5
##   NULL                                       2    94         1
##   South East                                 0     6        10
##   South West                                 0    39         3
##   West Midlands                              0     2         1
##   Yorkshire and The Humber                   0     7         3
##
##                            Photovoltaic Wind
##   East Midlands                    1546   72
##   East of England                  2507   75
##   London                            891    4
##   North East                        320   34
##   North West                        916   79
##   NULL                             1240  466
##   South East                       3943   42
##   South West                       3452  149
##   West Midlands                    1192   46
##   Yorkshire and The Humber         2397  144
```

```
#Converting from Table to Data Frame format
DF.ByRegion.Tech <- as.data.frame.matrix(ByRegion.Tech)
```

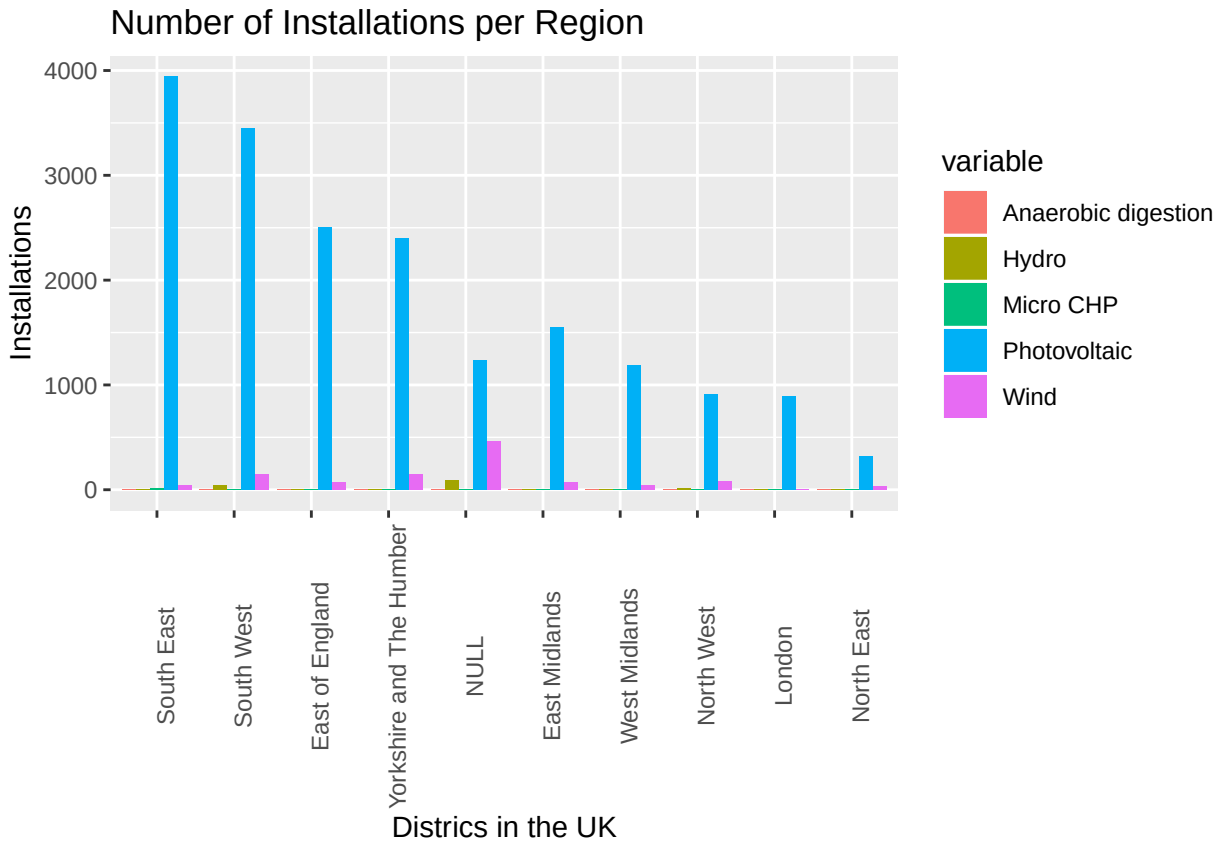
```
#Extractin Row Names and adding to Data Frame as First Column
DF.ByRegion.Tech <- setDT(DF.ByRegion.Tech, keep.rownames = "Region")[]
DF.ByRegion.Tech
```

```
##                      Region Anaerobic digestion Hydro Micro CHP
##  1:            East Midlands                   0     7         2
##  2:          East of England                   0     6         4
##  3:                   London                   0     0         1
##  4:               North East                   0     4         0
##  5:               North West                   0    11         5
##  6:                     NULL                   2    94         1
##  7:               South East                   0     6        10
##  8:               South West                   0    39         3
##  9:            West Midlands                   0     2         1
## 10: Yorkshire and The Humber                   0     7         3
##     Photovoltaic Wind
##  1:         1546   72
##  2:         2507   75
##  3:          891    4
##  4:          320   34
##  5:          916   79
##  6:         1240  466
##  7:         3943   42
##  8:         3452  149
##  9:         1192   46
## 10:         2397  144
```

How are the Renewable Generators installed in a per Region basis in the UK?

```
dfm.ByRegion.Tech <- melt(DF.ByRegion.Tech[,c("Region", "Anaerobic digestion", "Hydro", "Micro CHP", "Pl

#reorder(,-value) orders the bars from high to low.
ggplot(dfm.ByRegion.Tech,aes(x = reorder(Region, -value),y = value)) +
    geom_bar(aes(fill = variable),stat = "identity",position = "dodge") + theme(axis.text.x = element_te
```



Calculating Installed Generation Capacity per Region and per Technology Type

```
require(data.table)
DT <- data.table(UK.Renewables)
Region.Sums <- DT[ , .(Installed.Capacity = sum(InstalledCapacity)), by = .(GovernmentOfficeRegionName,
Region.Sums <- arrange(Region.Sums, GovernmentOfficeRegionName)
Region.Sums
```

```
##    GovernmentOfficeRegionName  TechnologyTypeName Installed.Capacity
## 1                East Midlands        Photovoltaic           4229.645
## 2                East Midlands                Wind            710.300
## 3                East Midlands               Hydro             93.500
## 4                East Midlands           Micro CHP              1.970
## 5             East of England        Photovoltaic           6017.649
## 6             East of England                Wind            590.800
## 7             East of England           Micro CHP              3.980
## 8             East of England               Hydro            545.000
## 9                      London        Photovoltaic           2368.862
```

```
## 10                    London                Wind            30.200
## 11                    London           Micro CHP             1.000
## 12                North East        Photovoltaic           866.848
## 13                North East                Wind           292.850
## 14                North East               Hydro            32.000
## 15                North West        Photovoltaic          2499.363
## 16                North West                Wind           498.500
## 17                North West               Hydro            73.200
## 18                North West           Micro CHP             4.940
## 19                      NULL        Photovoltaic          3734.267
## 20                      NULL                Wind          9166.100
## 21                      NULL               Hydro          7467.800
## 22                      NULL Anaerobic digestion           666.000
## 23                      NULL           Micro CHP             0.980
## 24                South East        Photovoltaic          9427.701
## 25                South East                Wind           248.892
## 26                South East           Micro CHP             9.900
## 27                South East               Hydro            78.800
## 28                South West        Photovoltaic          9323.920
## 29                South West               Hydro           457.600
## 30                South West                Wind          1000.700
## 31                South West           Micro CHP             3.090
## 32              West Midlands        Photovoltaic          3059.107
## 33              West Midlands                Wind           604.900
## 34              West Midlands               Hydro             8.500
## 35              West Midlands           Micro CHP             0.980
## 36   Yorkshire and The Humber        Photovoltaic          6820.464
## 37   Yorkshire and The Humber                Wind          1138.460
## 38   Yorkshire and The Humber               Hydro           285.260
## 39   Yorkshire and The Humber           Micro CHP             2.950
```

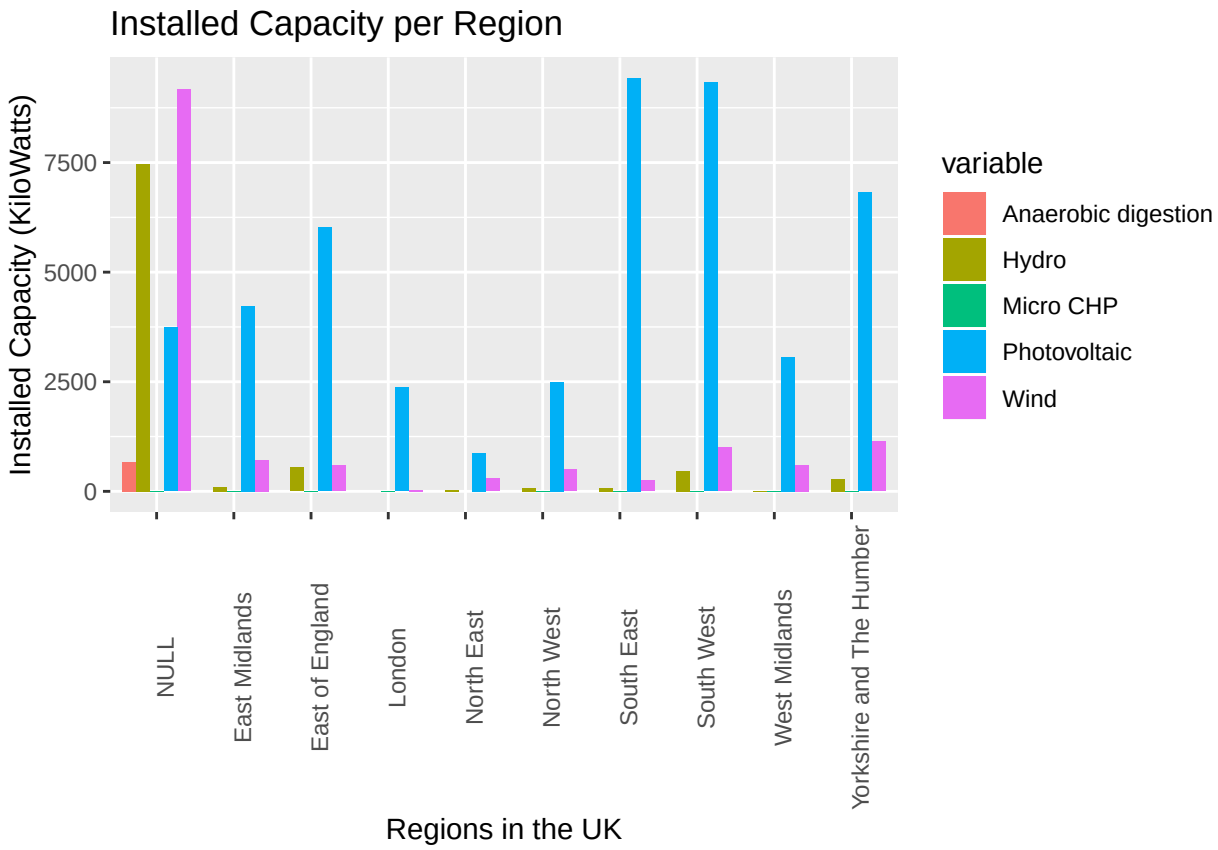
**Using Spread to Tyding up previous Data Frame.

```
spread.Regions.Sum <- spread(Region.Sums, TechnologyTypeName, Installed.Capacity)
```

```
dfm.Region.Sums <- melt(spread.Regions.Sum[,c("GovernmentOfficeRegionName", "Anaerobic digestion", "Hydr

#reorder(,-value) orders the bars from high to low.
ggplot(dfm.Region.Sums,aes(x = reorder(GovernmentOfficeRegionName, -value),y = value)) +
    geom_bar(aes(fill = variable),stat = "identity",position = "dodge") + theme(axis.text.x = element_te
```

```
## Warning: Removed 11 rows containing missing values (geom_bar).
```



We have seen from the study above that the most common installation type and most installed generation capacity both correspond to Photovoltaic technology.