Dams Package Demo

Gopi Goteti

November 3, 2016

This document provides examples on how to obtain data using the dams package and how to create summary graphics of the extracted data.

1 Data Attributes

If you have not already done so, load the package along with ggplot and maps (for graphics).

```
> require(dams)
> require(ggplot2)
> require(maps)
> require(mapproj)
```

2 Entire Dataset

Get the entire dataset. This might take a few moments.

```
> dams_all <- nid_cleaned
> dim(dams_all)

[1] 74096    61

> head(dams_all, 3)

                  Dam_Name  Other_Dam_Name  NID_ID Num_Separate_Struct Other_Structure_ID Lo
1        NIX MILL POND DAM    NIX MILL DAM AL00106                   0                     -
2 LIGHTSEY S MILL POND DAM   MILL POND DAM AL00533                   0                     -
3           ODUMS MILL DAM ODOMS MILL POND AL00890                   0                     -
  Latitude   County             River Owner_Type Private_Dam Dam_Designer Dam_Type Core Four
1  34.3483 FRANKLIN      EDGAR BRANCH Not Listed           N              Rockfill
2  32.9100     BIBB LIGHTSEY S BRANCH    Private           N                 Earth
3  33.9566   WALKER  BLACKWATER CREEK    Private           N               Gravity
  Primary_Purpose      All_Purposes Year_Completed Year_Modified Dam_Length Dam_Height
1           Other             Other           1800                        60         NA
2      Recreation        Recreation           1890                       350         NA
```

```
3           Other Recreation, Other           1850                       125         NA
  Structural_Height Hydraulic_Height NID_Height Max_Discharge Max_Storage Normal_Storage NID
1                25               25         25           140          55             55
2                13               10         13           250          80             80
3                12               12         12            NA         180            150
  Surface_Area Drainage_Area EAP Inspection_Date Inspection_Frequency Spillway_Type Spillway
1           NA            NA   N                                   0
2           NA            NA   N                                   0
3           NA            NA   N                                   0
  Outlet_Gates Volume Num_Locks Length_Locks Width_Locks Permitting_Authority Inspection_Aut
1                 NA        NA            0          NA                    N
2                 NA        NA            0          NA                    N
3                 NA        NA            0          NA                    N
  Enforcement_Authority Jurisdictional_Dam State_Reg_Dam State_Reg_Agency Fed_Funding Fed_De
1                     N                  N             N
2                     N                  N             N
3                     N                  N             N
  Fed_Construction Fed_Regulatory Fed_Inspection Fed_Operation Fed_Owner Fed_Other Source_Ag
1
2
3
  State  Submit_Date Url_Address           Congress_Rep Political_Party Congress_District
1    AL 01\\02\\2013          NA Robert B. Aderholt (R)               R              AL04
2    AL 01\\02\\2013          NA     Spencer Bachus (R)               R              AL06
3    AL 01\\02\\2013          NA Robert B. Aderholt (R)               R              AL04
```

3 Summary Graphics

Data for graphics.

```
> gfx_data <- dams_all[, c("Year_Completed", "State")]
> head(gfx_data)

  Year_Completed State
1           1800    AL
2           1890    AL
3           1850    AL
4           1880    AL
5           1881    AR
6           1877    CA
```

Counts of number of dams built per decade or other time period of interest.

```
> gfx_data$Year <- cut(gfx_data$Year_Completed,
+                      breaks = c(0, 1850, seq(1900, 2000, 10), 2014),
+                      labels = c("<1850", "1850-1900", "1910", "1920", "1930",
```

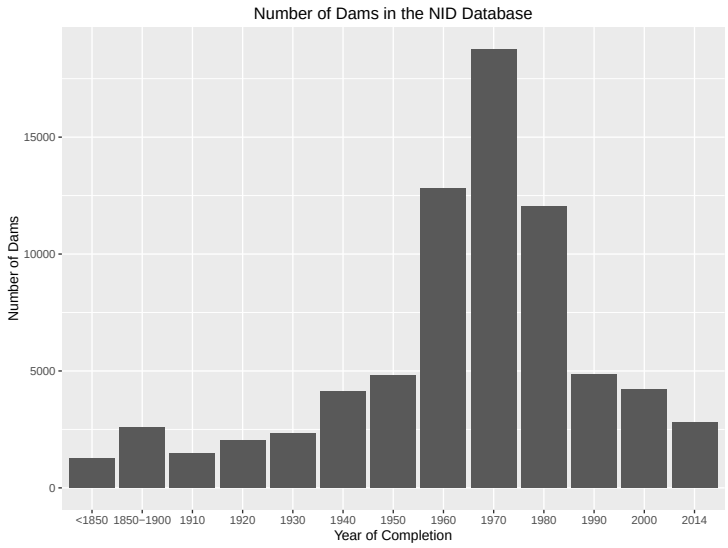

Figure 1: Number of Dams by Year of Completion

```
+                                   "1940", "1950", "1960", "1970", "1980", "1990",
+                                   "2000", "2014"))
> table(gfx_data$Year)

    <1850 1850-1900       1910       1920       1930       1940       1950       1960       1970
     1264       2588       1464       2015       2329       4130       4825      12816      18770
     1990       2000       2014
     4853       4221       2794

> year_counts <- as.data.frame(table(gfx_data$Year), stringsAsFactors = FALSE)
> colnames(year_counts) <- c("Year", "Count")
```

Histogram of number of dams by time period.

```
> gfx_bar <- ggplot(year_counts, aes(x = Year, y = Count))
> gfx_bar <- gfx_bar + geom_bar(position = "dodge", stat = "identity")
> gfx_bar <- gfx_bar + ylab("Number of Dams") + xlab("Year of Completion")
> gfx_bar <- gfx_bar + ggtitle("Number of Dams in the NID Database")

> plot(gfx_bar)
```

Counts of dams per state in the US mainland.

```
> gfx_data <- subset(gfx_data, !(State %in% c("AK", "HI", "PR", "GU")))
> sort(table(gfx_data$State))
```

```
  NV   DE   RI   MD   AZ   VT   IL   WV   ID   LA   NC   NM   NJ   NH   ME   UT   CT   FL
  43   51  174  305  319  358  391  413  431  441  444  453  510  590  597  617  700  781  7
  OR   IN   MI   MN   KY   TN   WI   AR   MA   OH   VA   PA   CA   WY   CO   NY   MS   AL
 839  847  914 1005 1025 1113 1116 1243 1244 1261 1307 1311 1500 1607 1704 1761 1981 2135 22
  SD   MT   GA   IA   OK   MO   KS   TX
2510 3236 3724 3913 4756 5002 5692 7253
```

Map of dams per state in the US mainland.

```
> state_counts <- as.data.frame(table(gfx_data$State), stringsAsFactors = FALSE)
> colnames(state_counts) <- c("state", "Count")
> # add long names of states
> state_names <- data.frame(state = state.abb,
+                           name = state.name,
+                           stringsAsFactors = FALSE)
> gfx_data <- merge(state_counts, state_names, by = "state")
> # change state name to lower case to be consistent with ggplot
> gfx_data$name <- tolower(gfx_data$name)
> # geo reference data on states from ggplot
> geo_state <- map_data("state")
> # merge data with above for graphics
> gfx_data <- merge(geo_state, gfx_data, by.x = "region", by.y = "name")
> gfx_data <- gfx_data[order(gfx_data$order), ]
> # discretize state counts
> color_breaks <- c(0, 100, 500, 1000, 2000, 3000, 4000, 5000, 7500)
> color_labels <- c("<100", "100 - 500", "500 - 1000", "1000 - 2000", "2000 - 3000",
+                   "3000 - 4000", "4000 - 5000", "5000 - 7500")
> gfx_data$dams <- cut(gfx_data$Count,
+                      breaks = color_breaks,
+                      labels = color_labels)
> gfx_map <- ggplot(data = gfx_data)
> gfx_map <- gfx_map + geom_polygon(aes(x = long, y = lat, group = group, fill = dams))
> gfx_map <- gfx_map + geom_path(data = geo_state, aes(x = long, y = lat, group = group,
+                                                      fill = NA))
> gfx_map <- gfx_map + labs(list(title = "Number of Dams in the NID Database",
+                                x = NULL, y = NULL))
> gfx_map <- gfx_map + guides(fill = guide_legend(title = "Number of Dams"))
> gfx_map <- gfx_map + scale_fill_brewer(palette = "Accent")
> gfx_map <- gfx_map + coord_map()

> plot(gfx_map)
```

4 Other Analyses: Flood Control Dams

A number of interesting analyses could be performed with the dataset. Of interest to water resources managers and hydrologists is the location of flood

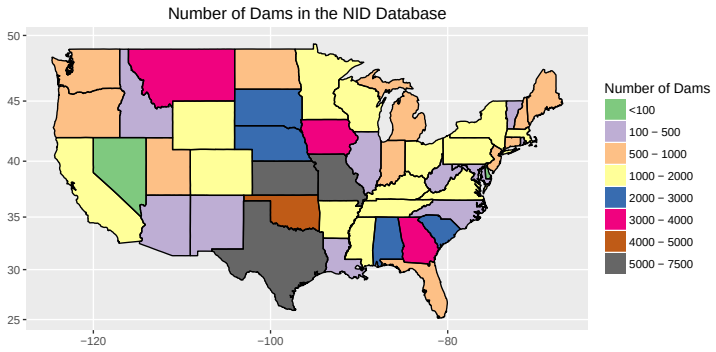

Figure 2: Number of Dams by State

control dams. It is interesting to see a few states like Texas have a large number of flood control dams.

```
> flood_dams <- subset(dams_all, Primary_Purpose == "Flood Control")
> table(flood_dams$State)

  AK   AL   AR   AZ   CA   CO   CT   DE   FL   GA   HI   IA   ID   IL   IN   KS   KY   LA
   4  131  221   91  190  230   20    2  171  369    6  369   12   49  147 1905  223   40
  ME   MI   MN   MO   MS   MT   NC   ND   NE   NH   NJ   NM   NV   NY   OH   OK   OR   PA
 113   28  305  897  596   98   79   91  950   43   33  208   12  139  112 2340   20  195
  SC   SD   TN   TX   UT   VA   VT   WA   WI   WV   WY
 112   89  209 2260  125  199   15  109  123  178  100
```