HUBZone Code – Re Run

John D. Harden

January 8, 2019

R Markdown

Anaylsis of the HUBZone program database for Washington, DC. Data was obtained/scraped from www.fpds.gov, https://usaspending.gov (https://usaspending.gov), HUBZone federal office. Data reveals an uneven distribution of dollars throughout the district, a notable finding considering that the point of the program was level the playing field for underserved District communities.

```
#Federal HUBZone Program Analysis
#loading libraries
library (tidyverse)
library (reshape)
library(knitr)
library(rmarkdown)
library(lubridate)

# loading data
HUBZoneRaw <- read_csv("HUBZone_Raw.csv",
                       col_types = cols(action_date = col_date(format = "%m/%d/%Y")))
options(stringsAsFactors = FALSE)
```

The sum of all the federal dollars secured in DC since the HUBZone program was implemented.

```
sum(HUBZoneRaw$federal_action_obligation)
```

```
## [1] 1118585908
```

Analyzing the top 11 firms using parent duns – names of firms are messy. Top 11 firms secured nearly $800 million over the course of the program.

```
# parse year
HUBZoneRaw$HUB_Year <- format(as.Date(HUBZoneRaw$action_date), "%Y")

firms_total <- group_by(HUBZoneRaw, recipient_parent_duns) %>%
  summarise(total_contracts = sum(federal_action_obligation)) %>%
  arrange(desc(total_contracts))

top11 <- head(firms_total, 11)

top11
```

```
## # A tibble: 11 x 2
##    recipient_parent_duns total_contracts
##    <chr>                           <dbl>
##  1 '605928246'                198334480.
##  2 '43227326'                 137804577.
##  3 '621796044'                120608845.
##  4 '796830201'                113581838.
##  5 '103312802'                 73272203.
##  6 '70899393'                  42851414.
##  7 '869308288'                 32183431.
##  8 '867994600'                 24750642.
##  9 '1627124'                   24623078.
## 10 '168847536'                 16171655.
## 11 '153637491'                 14324633.
```

Top 11 firms have secured about 70 percent of the total federal dollars awarded.

```
sum(top11$total_contracts) / sum(firms_total$total_contracts)
```

```
## [1] 0.7138538
```

Ward analysis – Wards one, four, five, seven and eight contain some of the most underserved communities in the nation's capital, accounting for more than 80 percent of the city's HUBZones. But those five wards have won less than 30 percent of the federal dollars awarded through HUBZone contracts.

```
#group by ward
ward_amount <- group_by(HUBZoneRaw, WARD) %>%
   summarise(total_contracts = sum(federal_action_obligation )) %>%
   arrange(desc(total_contracts))

ward_amount
```

```
## # A tibble: 8 x 2
##   WARD  total_contracts
##   <chr>           <dbl>
## 1 '2'        594766606.
## 2 '6'        187785736.
## 3 '4'        141786687.
## 4 '5'         83881376.
## 5 '8'         48822163.
## 6 '1'         22880459.
## 7 '7'         22442227.
## 8 '3'         16220655.
```

```
#percentage of awards by ward
ward_amount$total_contracts / sum(ward_amount$total_contracts)
```

```
## [1] 0.53171294 0.16787779 0.12675530 0.07498877 0.04364632 0.02045481
## [7] 0.02006303 0.01450104
```

```
#percentage of awards by ward

ward_amount <- group_by(HUBZoneRaw, WARD, HUB_Year) %>%
    summarise(total_contracts = sum(federal_action_obligation )) %>%
    arrange(desc(total_contracts))

ggplot(data = ward_amount) +
    geom_point(mapping = aes (x = WARD, y = total_contracts)) +
    facet_wrap ( ~ HUB_Year , nrow = 3 )
```