

Construction Industry in Puerto Rico - Data Summaries

Jack Keller (PR Studio 2024, Group 3)

2024-02-12

Overview

This document includes summaries and plots of the **Selected Statistics of the Construction Industry in Puerto Rico** dataset, provided by the *Junta de Planificación de Puerto Rico (Planning Board of Puerto Rico)*.

The data is from 2022.

Value of Construction Activity (Fiscal Semester)

Setup

```
knitr::opts_chunk$set(echo = TRUE)
library(tidyverse)
```

```
## — Attaching core tidyverse packages — tidyverse 2.0.0 —
## ✓ dplyr      1.1.4      ✓ readr      2.1.5
## ✓ forcats    1.0.0      ✓ stringr    1.5.0
## ✓ ggplot2    3.4.4      ✓ tibble     3.2.1
## ✓ lubridate  1.9.3      ✓ tidyr      1.3.0
## ✓ purrr      1.0.2
## — Conflicts — tidyverse_conflicts() —
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(readxl)
library(ggplot2)
library(conflicted)
# resolve conflicts with base R packages
conflict_prefer("filter", "dplyr")
```

```
## [conflicted] Will prefer dplyr::filter over any other package.
```

```
conflict_prefer("lag", "dplyr")
```

```
## [conflicted] Will prefer dplyr::lag over any other package.
```

```
# import data
# Value of Construction Activity (Fiscal Semesters)
cvfs <- read_excel("../Datasets/Junta de Planificacion/Tablas-de-las-Estadisticas-Seleccionadas-en-la-Industria-d
e-la-Construccion-2022.xlsx", sheet="T-2")
```

```
## New names:
## • `` -> `...2`
## • `` -> `...3`
## • `` -> `...4`
## • `` -> `...5`
## • `` -> `...6`
## • `` -> `...7`
## • `` -> `...8`
## • `` -> `...9`
## • `` -> `...10`
## • `` -> `...11`
## • `` -> `...12`
## • `` -> `...13`
## • `` -> `...14`
## • `` -> `...15`
## • `` -> `...16`
## • `` -> `...17`
## • `` -> `...18`
## • `` -> `...19`
## • `` -> `...20`
## • `` -> `...21`
## • `` -> `...22`
```

```
#View(cvfs)
```

Cleaning

```
# remove fully NA columns
cvfs <- select(cvfs, c(-1, -17:-21))

# transpose data
cvfs <- t(cvfs)

# convert cvfs matrix to tibble
cvfs <- cvfs %>%
  as_tibble()
```

```
## Warning: The `x` argument of `as_tibble.matrix()` must have unique column names if
## `.name_repair` is omitted as of tibble 2.0.0.
## i Using compatibility `.name_repair`.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```

# fix column names
colnames(cvfs)<- cvfs[16,]
cvfs <- cvfs[-16,]

# remove any columns with NA values
cvfs <- cvfs[, colSums(is.na(cvfs))==0]

# rename first column
colnames(cvfs)[1] <- 'fiscal_sem'

# split table in two, then join back together later
temp <- cvfs[, 38:ncol(cvfs)]
colnames(temp)[1] <- 'fiscal_sem' # set first column to same name as cvfs
cvfs <- cvfs[, 1:37]

# remove public data from cvfs, temp
cvfs <- cvfs[, 1:21]
temp <- temp[, 1:21]

# rename "Paid Claims" columns to match
colnames(cvfs)[ncol(cvfs)-2] <- 'paid_claims'
colnames(temp)[ncol(temp)-2] <- 'paid_claims'

# now, binding will work correctly since all column names match

# combine tibbles back together (all fiscal semesters will be organized in the first column now)
cvfs <- bind_rows(cvfs, temp)
rm(temp)

# rename all columns to be lowercase (snake_case)
colnames(cvfs) <- c('fiscal_sem', 'total', 'private', 'housing', 'urbanization', 'rural_proj', 'proj_under_500k',
'apt_building', 'commercial_and_industrial_building', 'hotel', 'hotel_villa', 'commercial', 'industrial', 'instal
lation_proj', 'pharma', 'telecomm', 'gas_station', 'renewable_nergy', 'paid_laims', 'property_insurance_housing',
'property_insurance_commercial')

# convert all columns except first to numeric type
cvfs[,2:ncol(cvfs)] <- cvfs[,2:ncol(cvfs)] %>% mutate_if(is.character, as.numeric)

```

```
## Warning: There were 3 warnings in `mutate()`.  
## The first warning was:  
## i In argument: `paid_laims = .Primitive("as.double")(paid_laims)`.  
## Caused by warning:  
## ! NAs introduced by coercion  
## i Run `dplyr::last_dplyr_warnings()` to see the 2 remaining warnings.
```

```
# dataset is cleaned  
# NOTE: paid_claims, property_insurance_housing, property_insurance_commercial values are missing for fiscal_sem  
J-2012 through 2017  
  
#view(cvfs)
```

Plots

```
data <- cvfs[-c(3,6,9,12,15,18,21,24,27,30), ] # exclude whole years

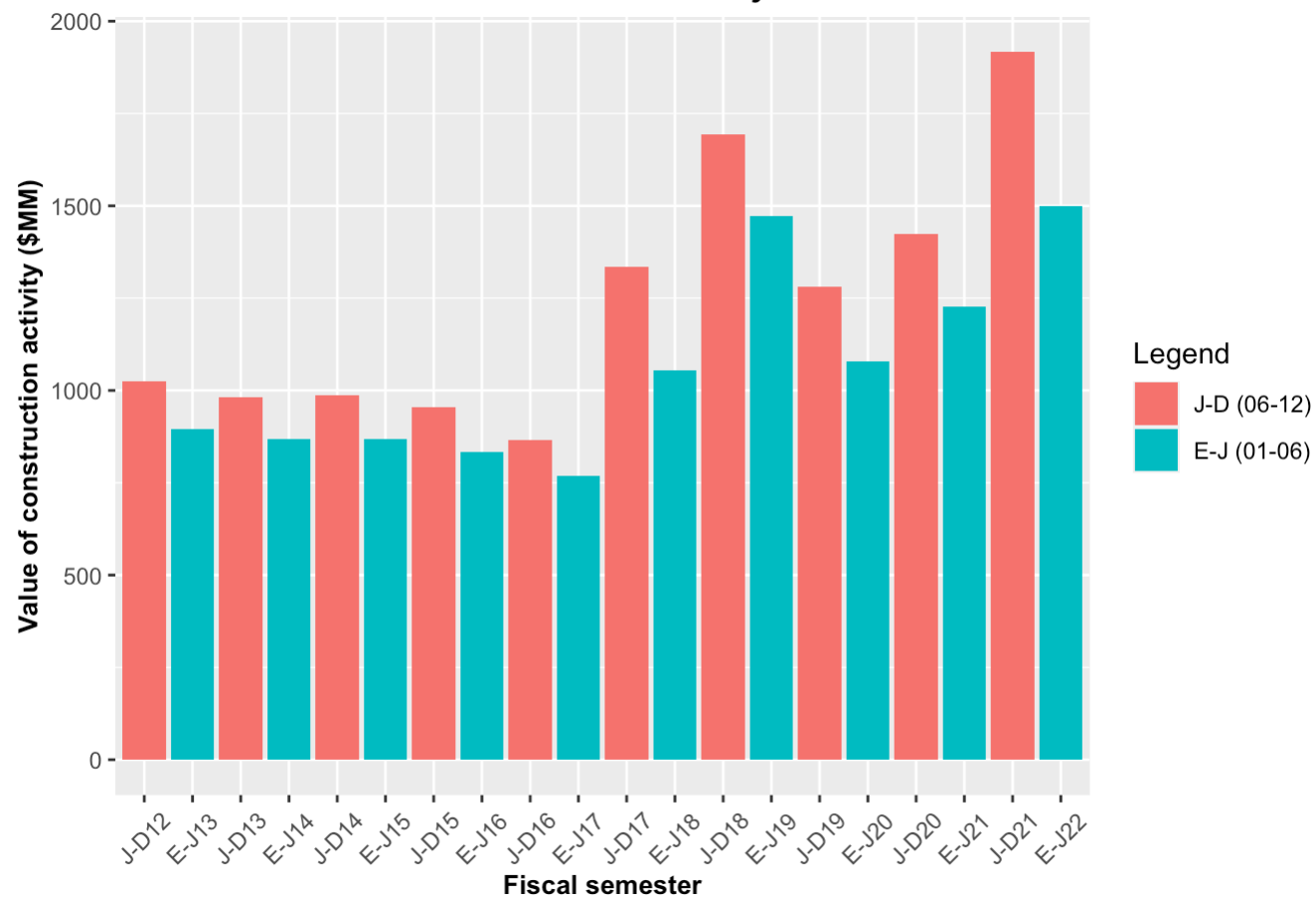
# Convert 'fiscal_sem' column to character vector
data$fiscal_sem <- as.character(data$fiscal_sem)

# Reorder it back into a factor with the original levels
data$fiscal_sem <- factor(data$fiscal_sem, levels = unique(data$fiscal_sem))

# ggplot(data, aes(x=fiscal_sem, y=private)) +
#   geom_line( color='blue', linewidth=0.5, linetype=1, group=1) +
#   theme(axis.text.x = element_text(angle = 45, vjust=0.5)) +
#   ggtitle("Total Value of Private Construction Activity.") +
#   xlab("Fiscal semester") +
#   ylab("Value of construction ($ in millions)") +
#   theme(
#     plot.title = element_text(size = 12, face = "bold.italic"),
#     axis.title.x = element_text(size = 10, face = "bold"),
#     axis.title.y = element_text( size = 10, face = "bold"))

ggplot(data, aes(x=fiscal_sem, y=private, fill=(as.numeric(fiscal_sem) %% 2 == 0))) +
  geom_bar(stat = "identity") +
  scale_fill_discrete(labels=c('J-D (06-12)', 'E-J (01-06)'), name='Legend') +
  theme(axis.text.x = element_text(angle = 45, vjust=0.5)) +
  ggtitle("Total Value of Private Construction Activity in Puerto Rico") +
  xlab("Fiscal semester") +
  ylab("Value of construction activity ($MM)") +
  theme(
    plot.title = element_text(size = 12, face = "bold.italic"),
    axis.title.x = element_text(size = 10, face = "bold"),
    axis.title.y = element_text( size = 10, face = "bold"))
```

Total Value of Private Construction Activity in Puerto Rico



```
data <- cvfs[c(3,6,9,12,15,18,21,24,27,30), -c(2, 3)] # whole years only
data
```



```
## # A tibble: 10 × 19
##   fiscal_sem housing urbanization rural_proj proj_under_500k apt_building
##   <chr>         <dbl>         <dbl>         <dbl>         <dbl>         <dbl>
## 1 2013          630.          238.           0           120.          272.
## 2 2014          547.          206.           0           105.          236.
## 3 2015          571.          232.           0            95.5          244.
## 4 2016          575.          233.           0           100.          242.
## 5 2017          480.          188.           0           103.          189.
## 6 2018          469.          134.           0           139.          196.
## 7 2019          506.          117.           0           201.          188.
## 8 2020          478.          146.           0           141.          191.
## 9 2021p          465.          161.           0           111.          193.
## 10 2022p          717.          228.           0           231.          259.
## # i 13 more variables: commercial_and_industrial_building <dbl>, hotel <dbl>,
## # hotel_villa <dbl>, commercial <dbl>, industrial <dbl>,
## # installation_proj <dbl>, pharma <dbl>, telecomm <dbl>, gas_station <dbl>,
## # renewable_nergy <dbl>, paid_laims <dbl>, property_insurance_housing <dbl>,
## # property_insurance_commercial <dbl>
```