Analysis of building permits and school data

Mehdi Khan

December 13, 2017

Introduction:

Building permit and school data from 2002 to 2017 have been collected to see if there is any significant relationship between the building permits approved by the county and the changes in the schools' percentage capacity (enrollment/capacity) and/or enrollment in schools. Only the completed building permits were considered here.

Research question:

Does the number of completed building permits have any impact on the percentage capacity and/or enrollment in the schools?

Result summary:

No significant relationship was found. Very weak correlations were found between building permits and other two variables (permit, percentage capacity).

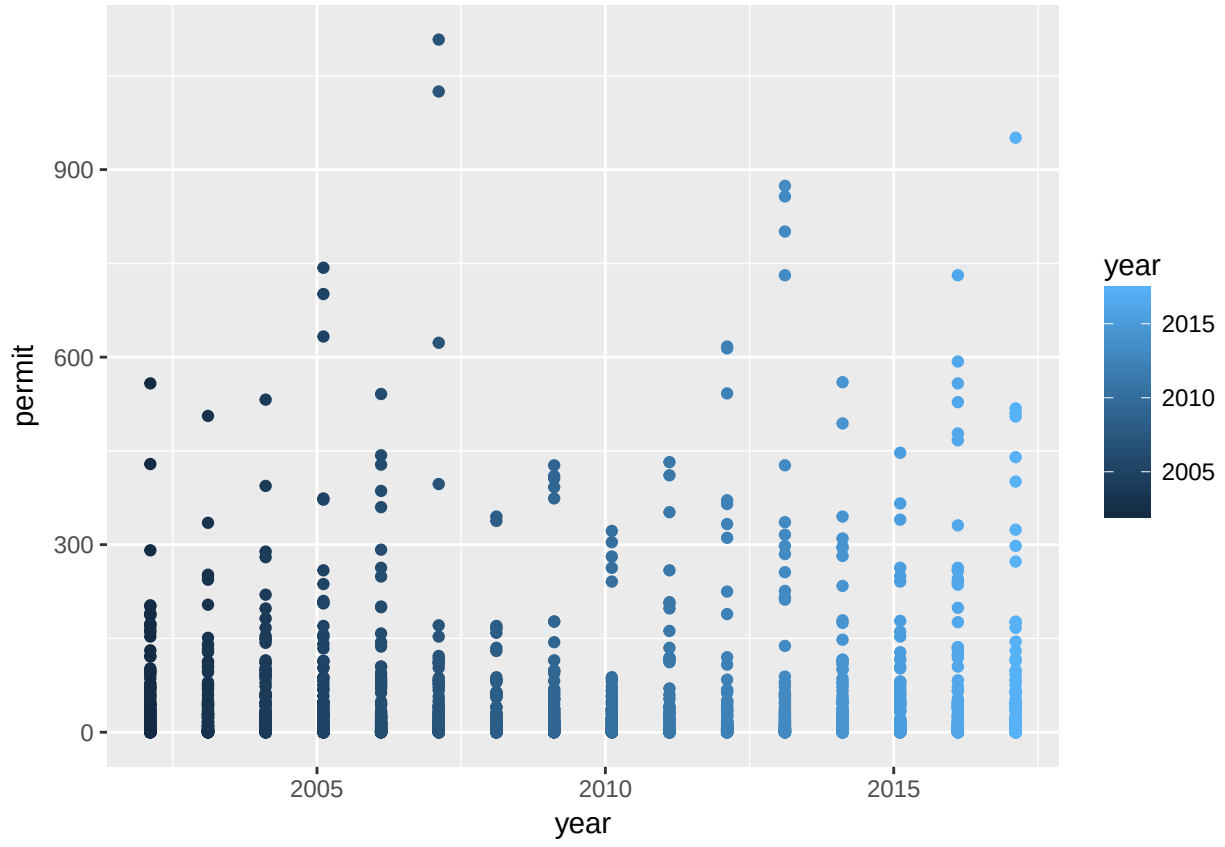
Analysis:

```
school_info <- read.csv("newSchoolEnrollment.csv", sep = ",", stringsAsFactors = FALSE,
    header = FALSE)
fields <- as.character(as.vector(school_info[2, ]))
colnames(school_info) <- fields
schoolDS_new <- school_info[-c(1:2, 113), -c(20, 53, 54)]
schoolDS_new <- rename(schoolDS_new, school = "")
```

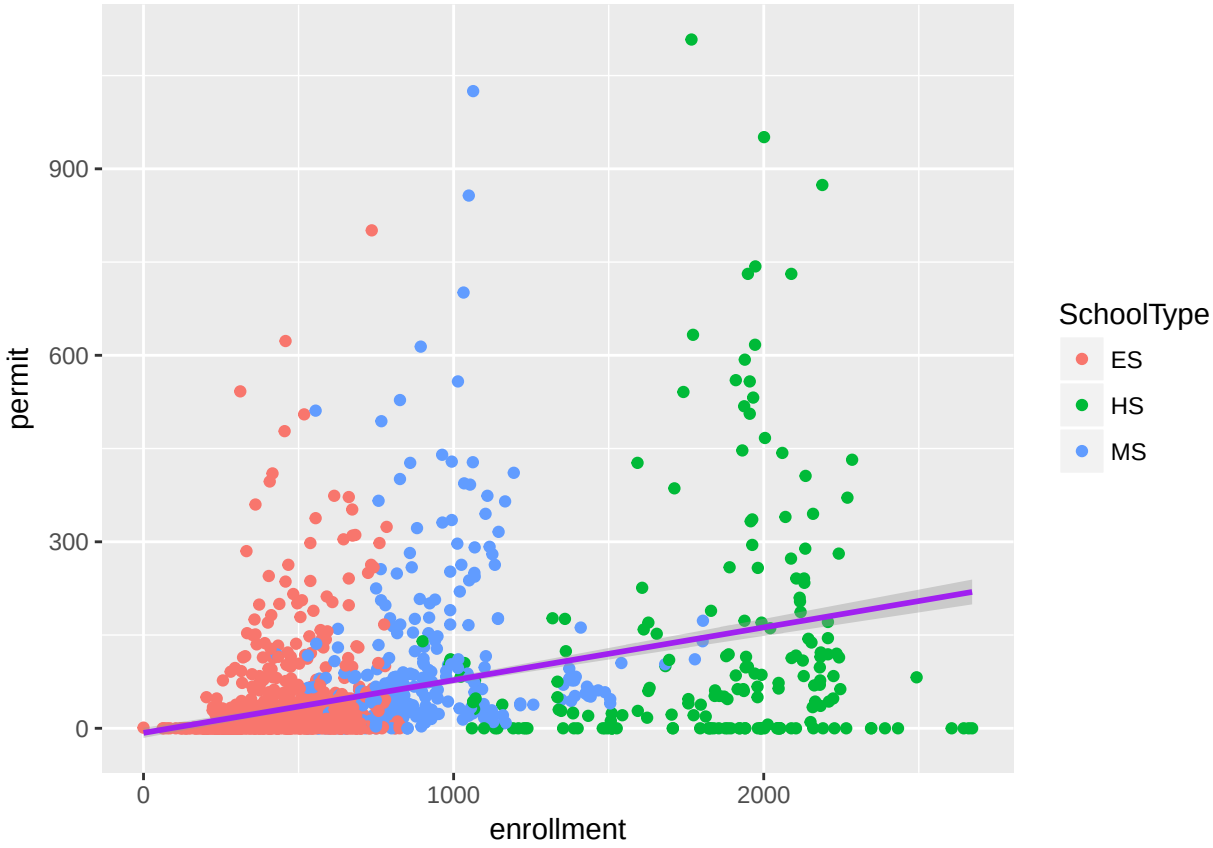
```
school_new <- gather(schoolDS_new, key, value, -c(1:3))
year <- school_new$key


school_new[grepl("\\.", school_new$key) == FALSE, "key"] <- "enrollment"
school_new[grepl("\\.1", school_new$key) == TRUE, "key"] <- "capacity"
school_new[grepl("\\.2", school_new$key) == TRUE, "key"] <- "%utilization"
school_new[grepl("\\.3", school_new$key) == TRUE, "key"] <- "permit"

school_new <- cbind(school_new, year)

enrollDs <- school_new[school_new$key == "enrollment", -4]
capDs <- school_new[school_new$key == "capacity", -c(1, 2, 4, 6)]
utilDs <- school_new[school_new$key == "%utilization", -c(1, 2, 4,
    6)]
permitDs <- school_new[school_new$key == "permit", -c(1, 2, 4, 6)]

enrollDs <- rename(enrollDs, enrollment = "value")
```

```r
capDs <- rename(capDs, capacity = "value")
utilDs <- rename(utilDs, `%utilization` = "value")
permitDs <- rename(permitDs, permit = "value")

schoolDF <- cbind(enrollDs, capDs, utilDs, permitDs)
schoolDF <- schoolDF[, -c(6, 8, 10)]
year <- schoolDF$year
schoolDF <- schoolDF[, -5]
schoolDF <- cbind(year, schoolDF)

schoolDF$enrollment <- as.numeric(schoolDF$enrollment)
schoolDF$capacity <- as.numeric(schoolDF$capacity)
schoolDF$permit <- as.numeric(as.character(schoolDF$permit))
# schoolDF$`%utilization` <- as.numeric(schoolDF$`%utilization`)
schoolDF$year <- as.Date(schoolDF$year, "%Y")

schoolDF <- mutate(schoolDF, percent_capacity = schoolDF$enrollment/schoolDF$capacity)
schoolDF <- na.omit(schoolDF)
# schoolDF <- schoolDF[-schoolDF$percent_capacity==Inf,]
schoolDF <- subset(schoolDF, !schoolDF$percent_capacity == Inf)
```

```r
p1 <- ggplot(schoolDF, aes(x = year, y = permit, col = year)) + geom_point()
p1
```




```r
ggplot(schoolDF, aes(x = enrollment, y = permit, col = SchoolType)) +
    geom_point() + geom_smooth(method = lm, col = "purple")
```



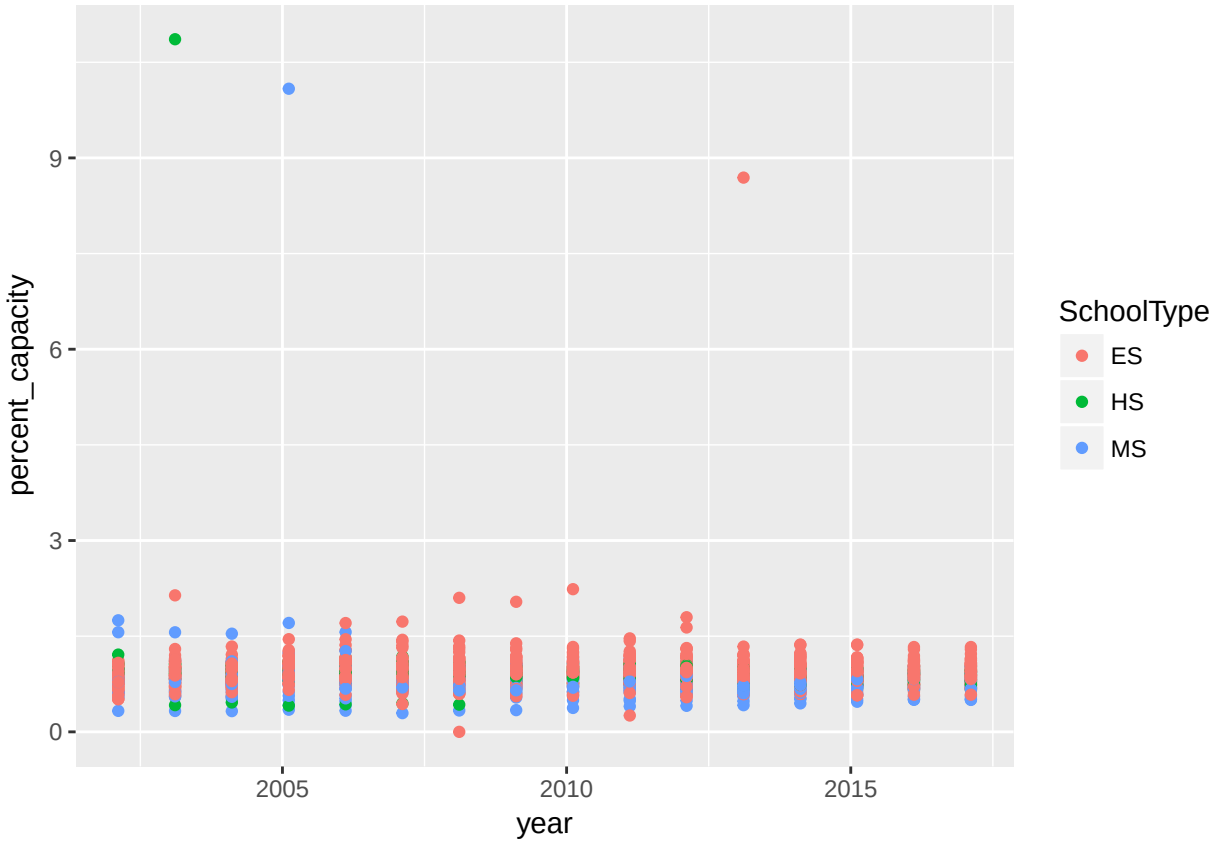

```
cor(schoolDF$permit, schoolDF$enrollment, method = c("pearson", "kendall",
    "spearman"))
```

```
## [1] 0.3790642
```

```
cor.test(schoolDF$permit, schoolDF$enrollment, method = c("pearson",
    "kendall", "spearman"))
```

```
##
##  Pearson's product-moment correlation
##
## data:  schoolDF$permit and schoolDF$enrollment
## t = 17.122, df = 1747, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.3382018 0.4185000
## sample estimates:
##       cor
## 0.3790642
```

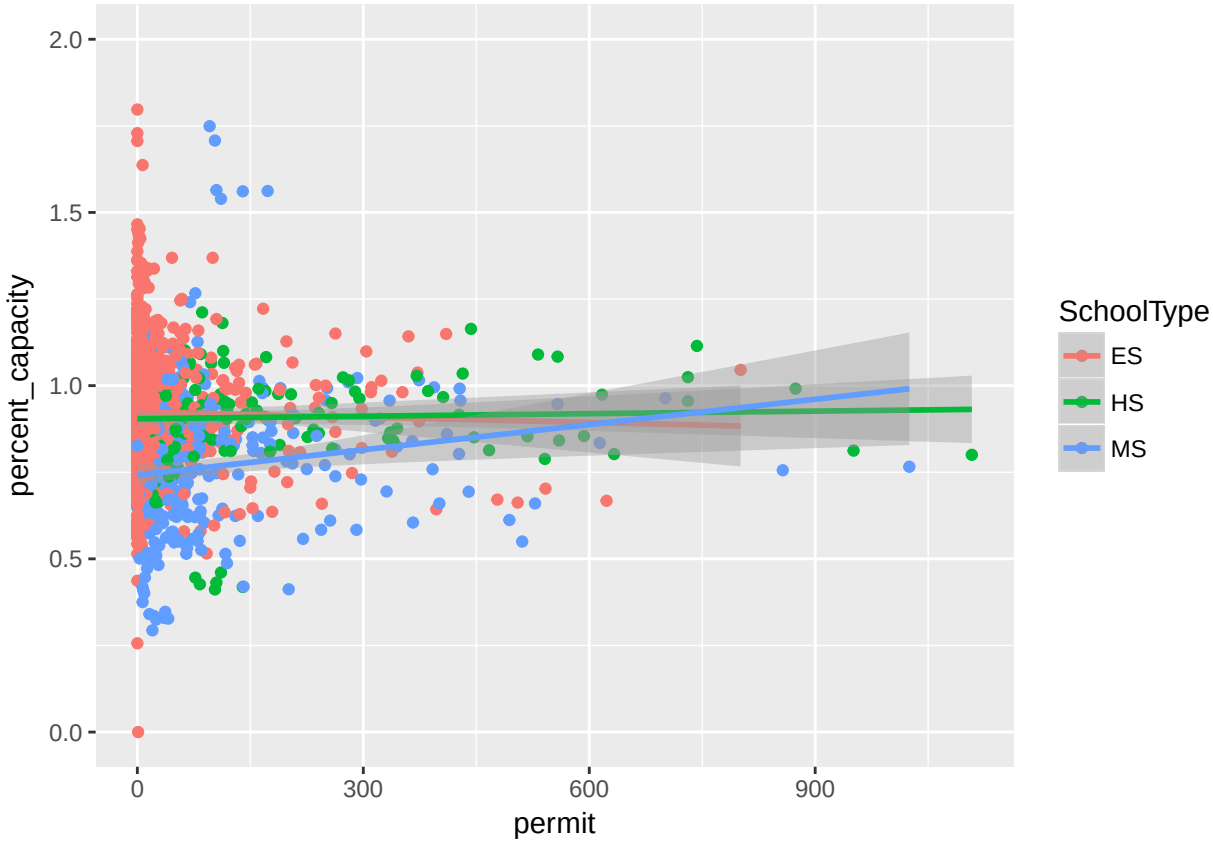
```
ggplot(schoolDF, aes(x = year, y = percent_capacity, col = SchoolType)) +
    geom_point()
```




```
ggplot(schoolDF, aes(x = permit, y = percent_capacity, col = SchoolType)) +
    geom_point() + ylim(0, 2) + geom_smooth(method = lm)
```

```
## Warning: Removed 7 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 7 rows containing missing values (geom_point).
```



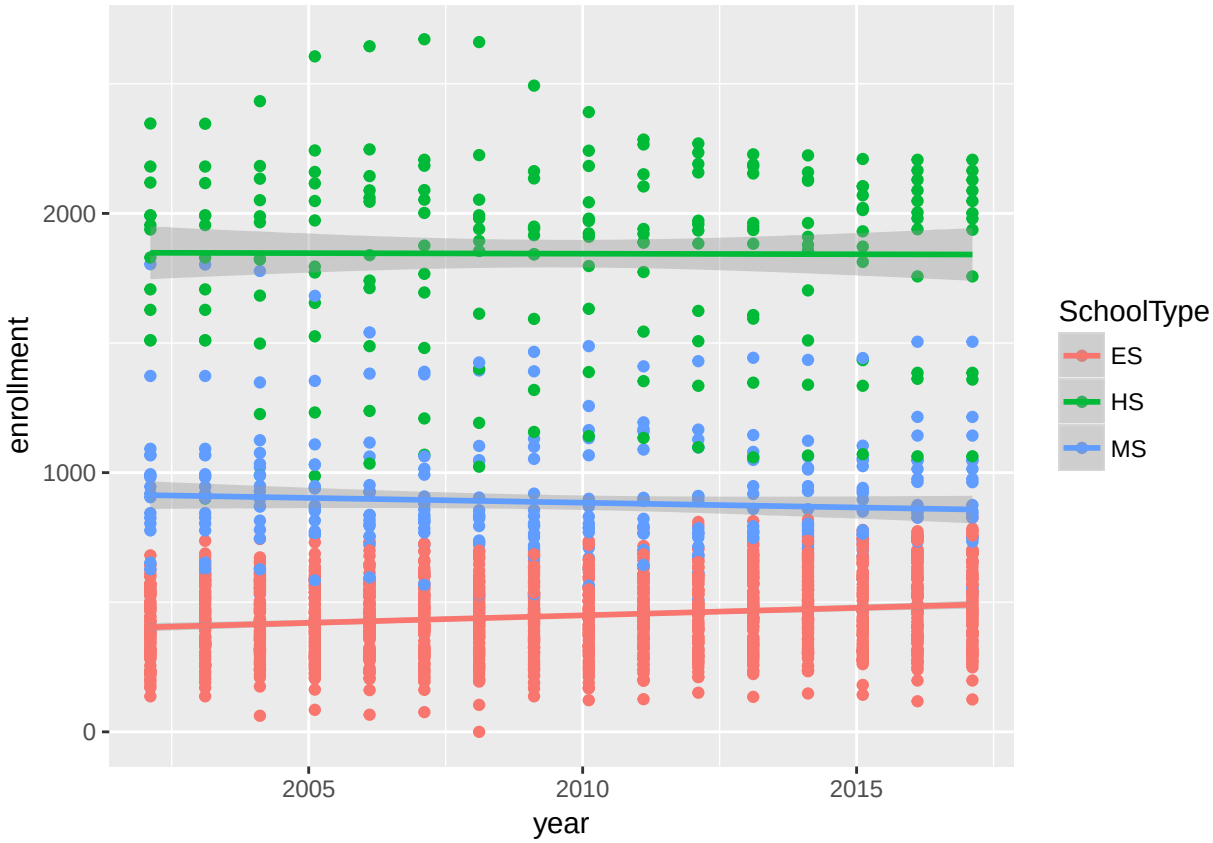

```
cor(schoolDF$permit, schoolDF$percent_capacity, method = c("pearson",
    "kendall", "spearman"))
```

```
## [1] 0.0520807
```

```
cor.test(schoolDF$permit, schoolDF$percent_capacity, method = c("pearson",
    "kendall", "spearman"))
```

```
##
##  Pearson's product-moment correlation
##
## data:  schoolDF$permit and schoolDF$percent_capacity
## t = 2.1798, df = 1747, p-value = 0.02941
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.005222081 0.098711106
## sample estimates:
##       cor
## 0.0520807
```

```
ggplot(schoolDF, aes(x = year, y = enrollment, col = SchoolType)) +
    geom_point() + geom_smooth(method = lm)
```




```
cor(schoolDF$permit, schoolDF$enrollment, method = c("pearson", "kendall",
    "spearman"))
```

```
## [1] 0.3790642
```

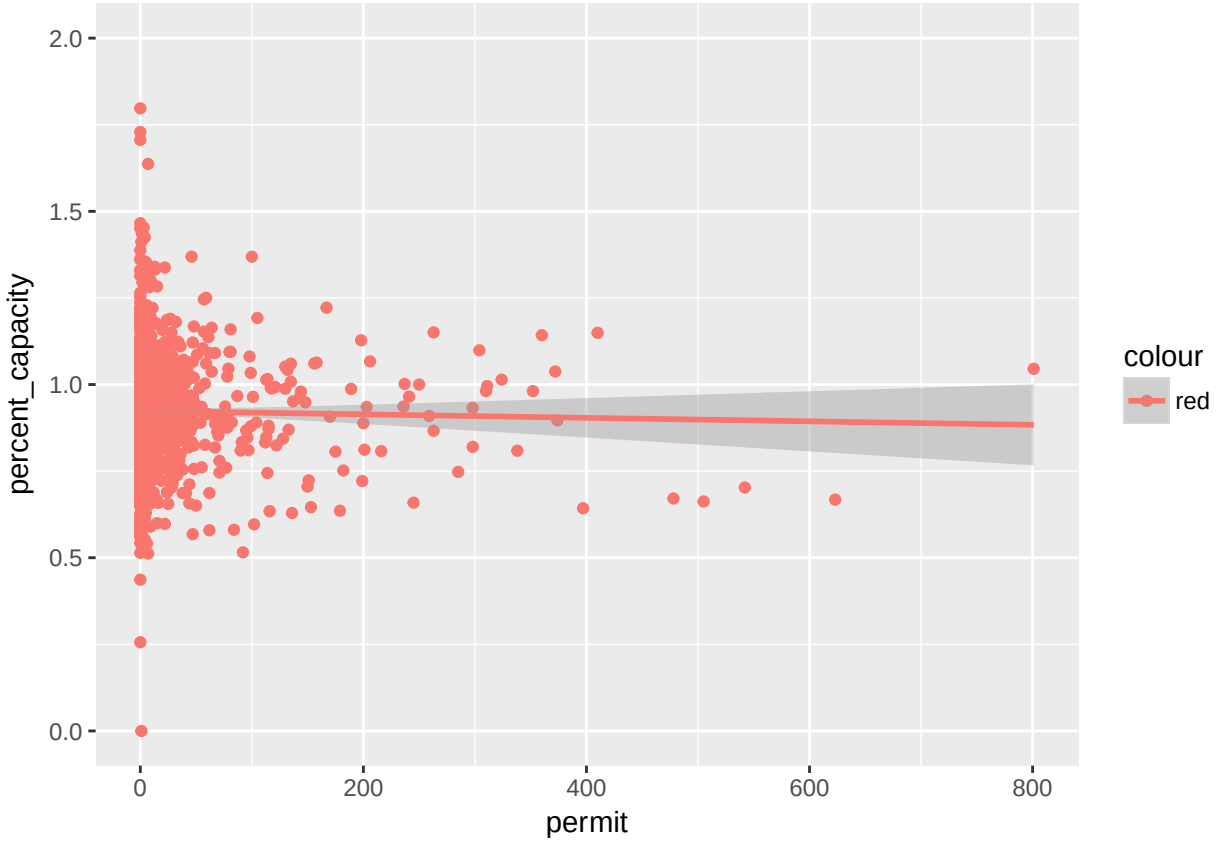
```
cor.test(schoolDF$permit, schoolDF$enrollment, method = c("pearson",
    "kendall", "spearman"))
```

```
##
##  Pearson's product-moment correlation
##
## data:  schoolDF$permit and schoolDF$enrollment
## t = 17.122, df = 1747, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.3382018 0.4185000
## sample estimates:
##       cor
## 0.3790642
```

Elementary Schools

```
schoolDFES <- schoolDF[schoolDF$SchoolType == "ES", ]
ggplot(schoolDFES, aes(x = permit, y = percent_capacity, col = "red")) +
    geom_point() + ylim(0, 2) + geom_smooth(method = lm)
```

```
## Warning: Removed 5 rows containing non-finite values (stat_smooth).
## Warning: Removed 5 rows containing missing values (geom_point).
```




```
cor(schoolDFES$permit, schoolDFES$percent_capacity, method = c("pearson",
    "kendall", "spearman"))
```

```
## [1] 0.04995587
```

```
cor.test(schoolDFES$permit, schoolDFES$percent_capacity, method = c("pearson",
    "kendall", "spearman"))
```

```
##
##  Pearson's product-moment correlation
##
## data:  schoolDFES$permit and schoolDFES$percent_capacity
## t = 1.7691, df = 1251, p-value = 0.07712
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.005438614  0.105044692
## sample estimates:
##        cor
## 0.04995587
```

```
# ggplot(schoolDFES,aes(x=enrollment, y=permit, col='red'))+
# geom_point()
```

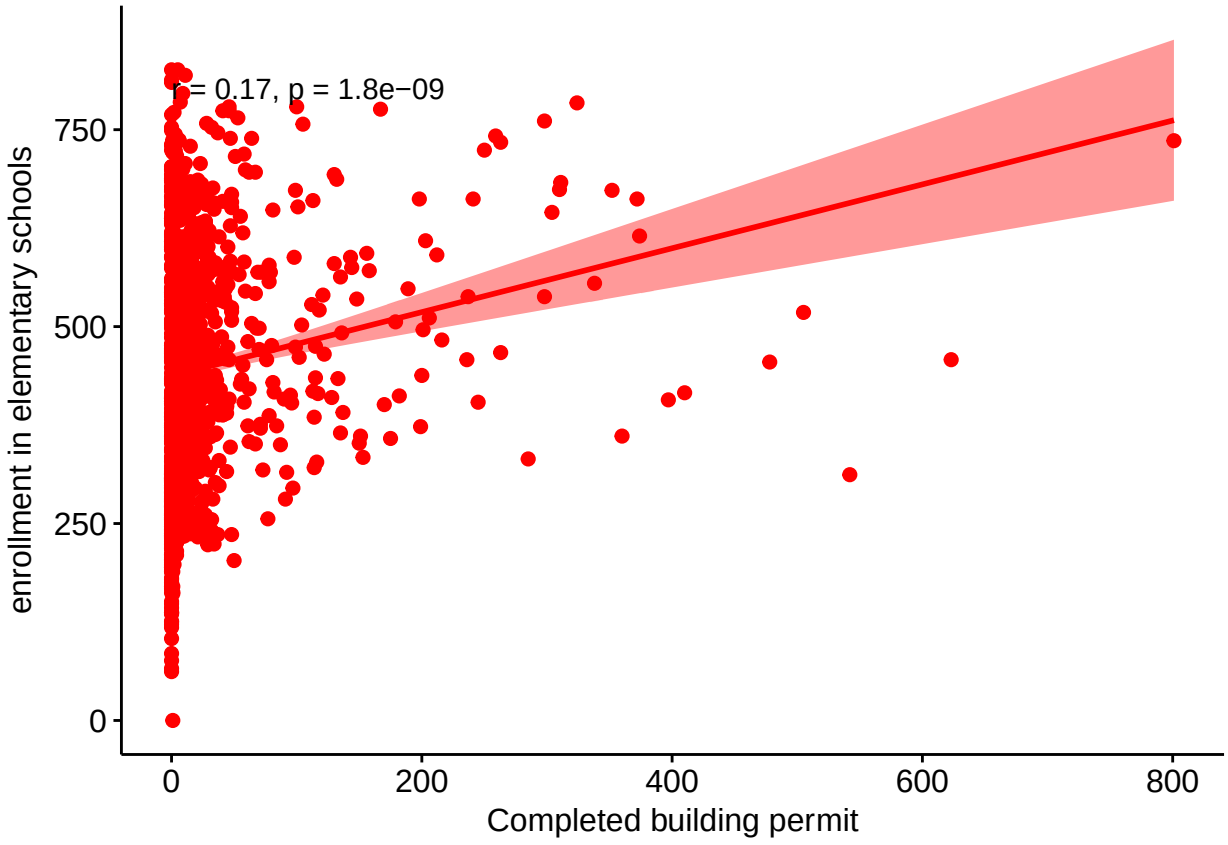
```
cor(schoolDFES$permit, schoolDFES$enrollment, method = c("pearson",
    "kendall", "spearman"))
```

```
## [1] 0.1687993
```

```
cor.test(schoolDFES$permit, schoolDFES$enrollment, method = c("pearson",
    "kendall", "spearman"))
```

```
##
##  Pearson's product-moment correlation
##
## data:  schoolDFES$permit and schoolDFES$enrollment
## t = 6.0573, df = 1251, p-value = 1.829e-09
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.1144901 0.2221025
## sample estimates:
##       cor
## 0.1687993
```

```
ggscatter(schoolDFES, x = "permit", y = "enrollment", add = "reg.line",
    conf.int = TRUE, cor.coef = TRUE, cor.method = "pearson", xlab = "Completed building permit",
    ylab = "enrollment in elementary schools", color = "red")
```

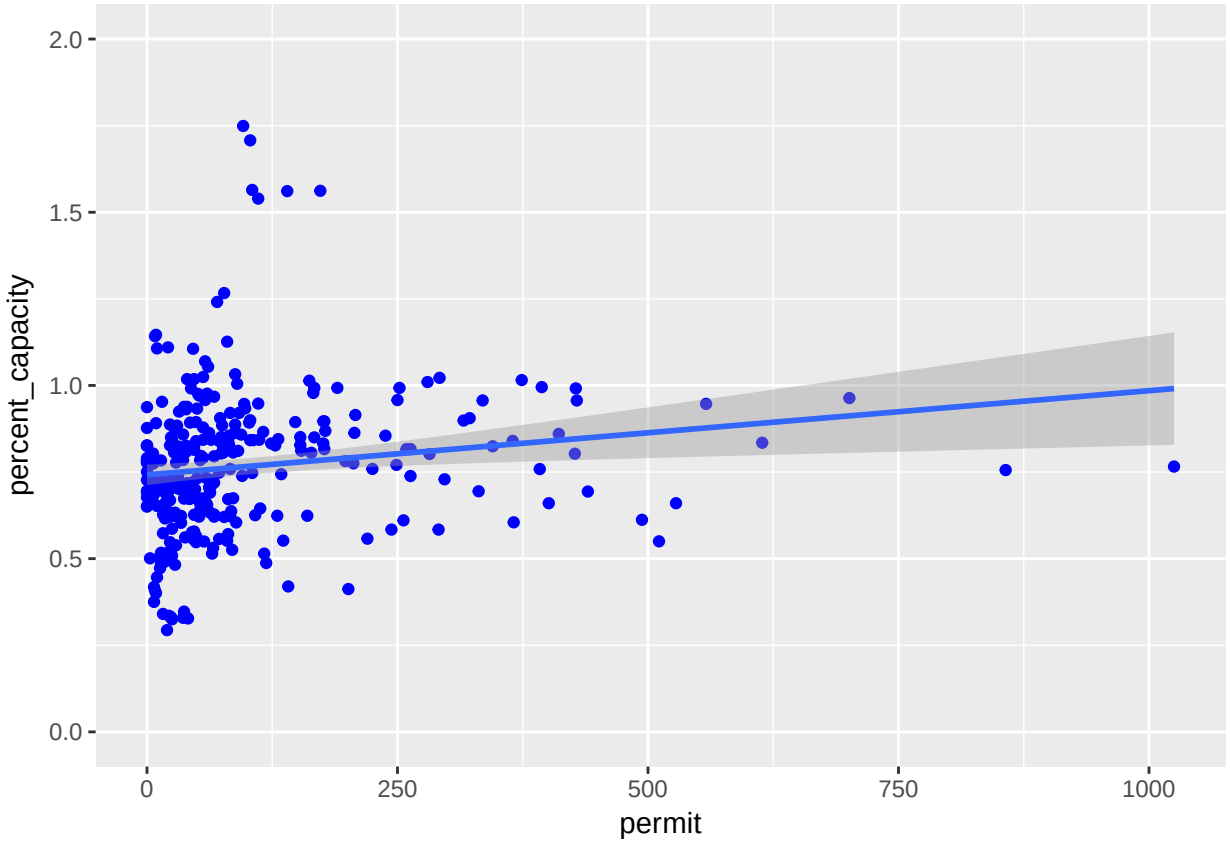


Middle Schools

```
schoolDFMS <- schoolDF[schoolDF$SchoolType == "MS", ]
ggplot(schoolDFMS, aes(x = permit, y = percent_capacity)) + geom_point(color = "blue") +
    ylim(0, 2) + geom_smooth(method = lm)
```

```
## Warning: Removed 1 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```



Correlation between enrollment and percentage capacity

```
cor(schoolDFMS$permit, schoolDFMS$percent_capacity, method = c("pearson",
    "kendall", "spearman"))
```

```
## [1] 0.03625019
```

```
cor.test(schoolDFMS$permit, schoolDFMS$percent_capacity, method = c("pearson",
    "kendall", "spearman"))
```

```
##
##  Pearson's product-moment correlation
##
## data:  schoolDFMS$permit and schoolDFMS$percent_capacity
## t = 0.63038, df = 302, p-value = 0.5289
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.07655429  0.14813840
```

```
## sample estimates:
##        cor
## 0.03625019
```

```
# ggplot(schoolDFMS,aes(x=enrollment, y=permit, col='blue'))+
# geom_point(color='blue')
```

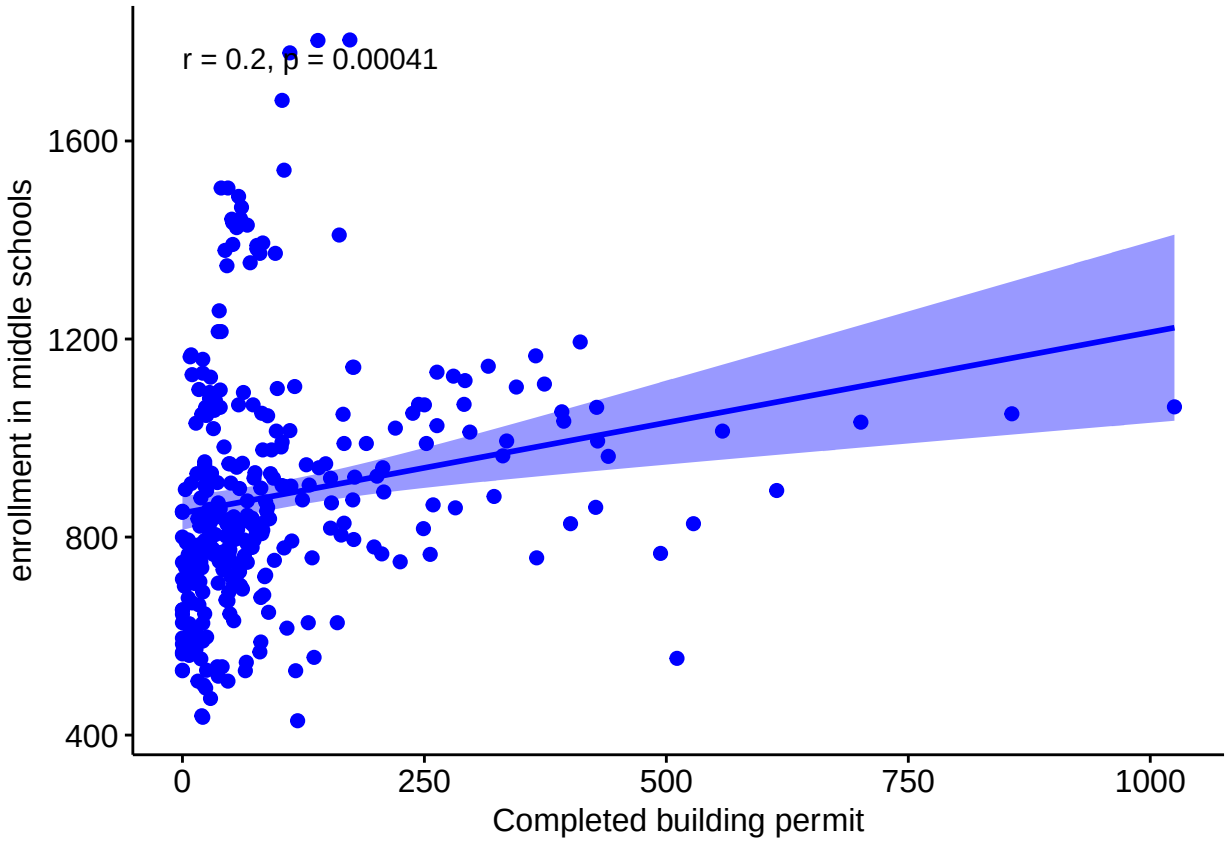
```
cor(schoolDFMS$permit, schoolDFMS$enrollment, method = c("pearson",
    "kendall", "spearman"))
```

```
## [1] 0.2012391
```

```
cor.test(schoolDFMS$permit, schoolDFMS$enrollment, method = c("pearson",
    "kendall", "spearman"))
```

```
##
##  Pearson's product-moment correlation
##
## data:  schoolDFMS$permit and schoolDFMS$enrollment
## t = 3.5702, df = 302, p-value = 0.0004148
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.09080236 0.30678640
## sample estimates:
##       cor
## 0.2012391
```

```
ggscatter(schoolDFMS, x = "permit", y = "enrollment", add = "reg.line",
    conf.int = TRUE, cor.coef = TRUE, cor.method = "pearson", xlab = "Completed building permit",
    ylab = "enrollment in middle schools", color = "blue")
```

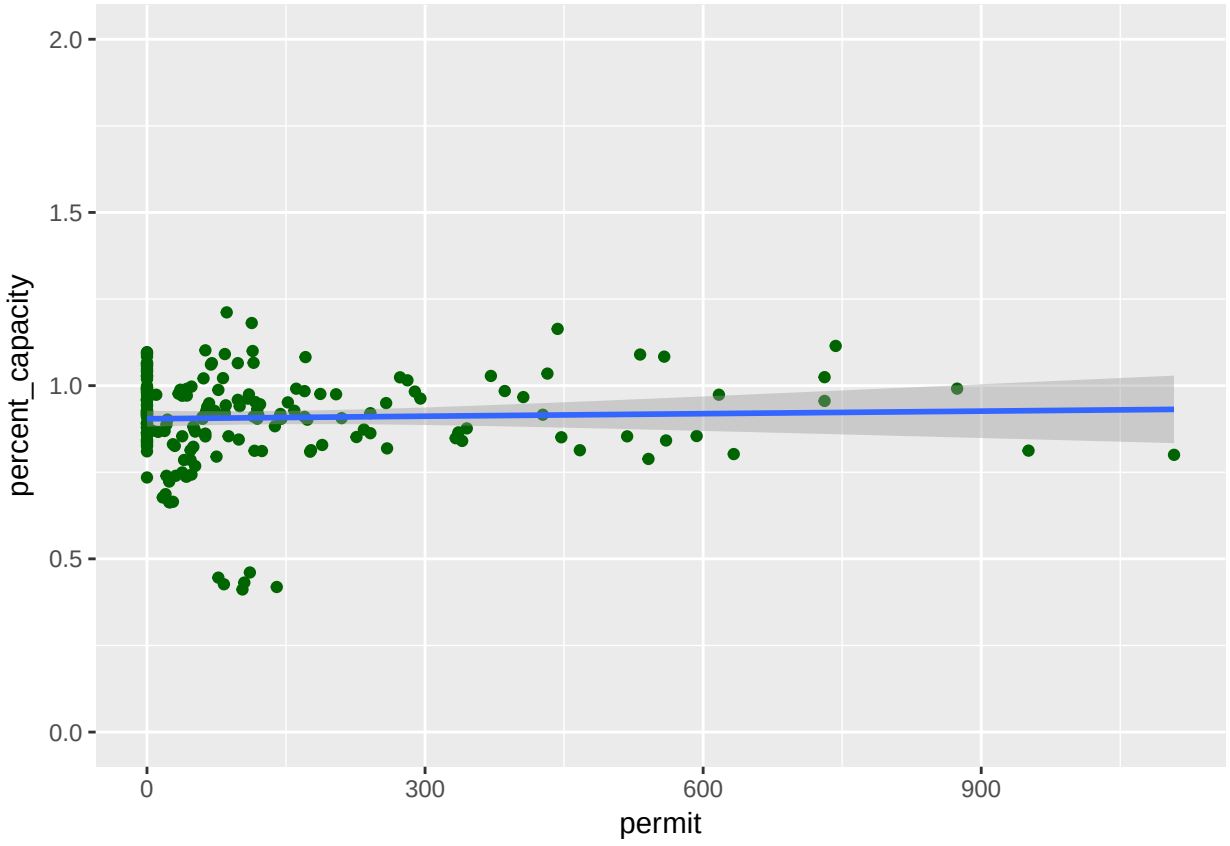


High Schools

```
schoolDFHS <- schoolDF[schoolDF$SchoolType == "HS", ]
ggplot(schoolDFHS, aes(x = permit, y = percent_capacity)) + geom_point(color = "darkgreen") +
    ylim(0, 2) + geom_smooth(method = lm)
```

```
## Warning: Removed 1 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```



Correlation between enrollment and percentage capacity

```
cor(schoolDFHS$permit, schoolDFHS$percent_capacity, method = c("pearson",
    "kendall", "spearman"))
```

```
## [1] 0.1410996
```

```
cor.test(schoolDFHS$permit, schoolDFHS$percent_capacity, method = c("pearson",
    "kendall", "spearman"))
```

```
##
##  Pearson's product-moment correlation
##
## data:  schoolDFHS$permit and schoolDFHS$percent_capacity
## t = 1.9646, df = 190, p-value = 0.05092
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.0005190177  0.2771698446
## sample estimates:
##       cor
## 0.1410996
```

```
# ggplot(schoolDFHS,aes(x=enrollment, y=permit))+
# geom_point(color='darkgreen')
```

Correlation between enrollment and building permit

```
cor(schoolDFHS$permit, schoolDFHS$enrollment, method = c("pearson",
    "kendall", "spearman"))
```

```
## [1] 0.1574705
```

```
cor.test(schoolDFHS$permit, schoolDFHS$enrollment, method = c("pearson",
    "kendall", "spearman"))
```

```
##
##  Pearson's product-moment correlation
##
## data:  schoolDFHS$permit and schoolDFHS$enrollment
## t = 2.198, df = 190, p-value = 0.02916
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.01622401 0.29255509
## sample estimates:
##       cor
## 0.1574705
```

```
ggscatter(schoolDFHS, x = "permit", y = "enrollment", add = "reg.line",
    conf.int = TRUE, cor.coef = TRUE, cor.method = "pearson", xlab = "Completed building permit",
    ylab = "enrollment in high schools", color = "darkgreen")
```