ProblemSet2_new

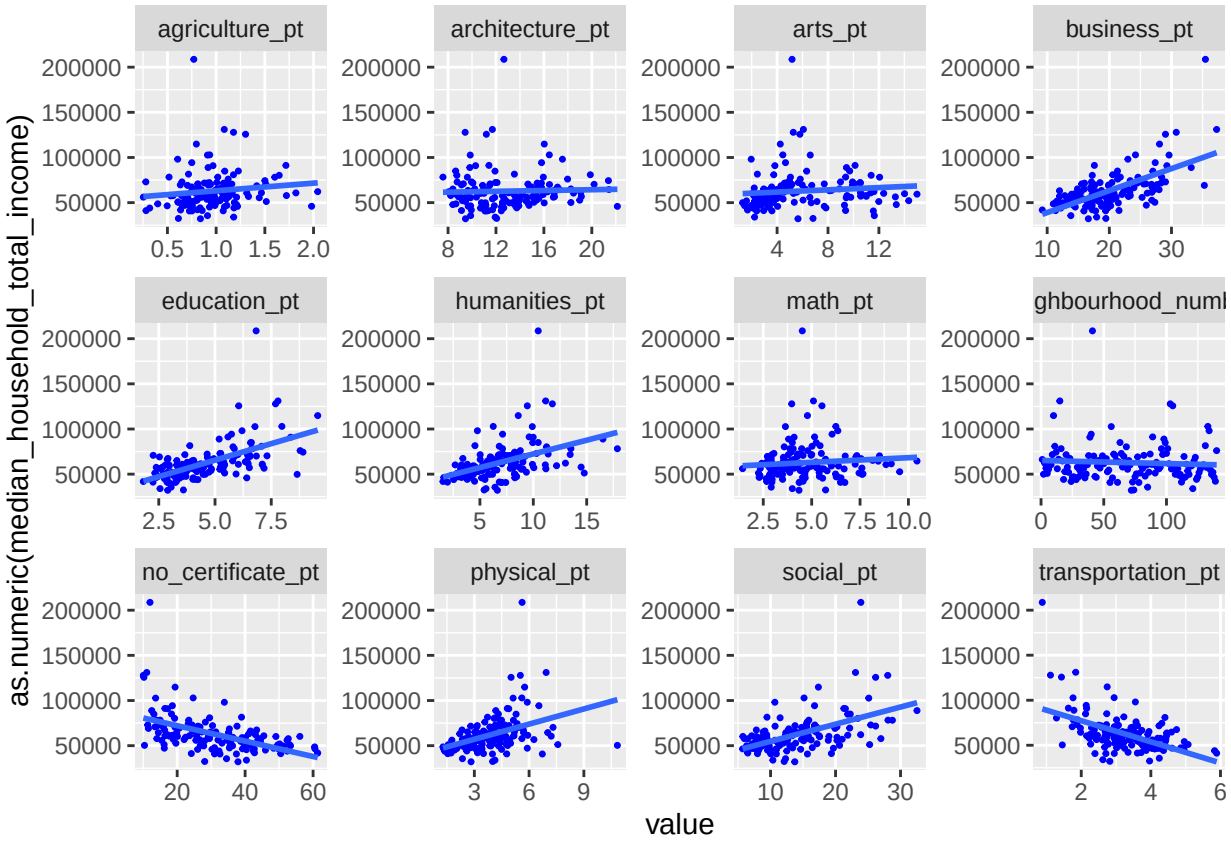
Ke-li Chiu & Diego Mamanche Castellanos

06/02/2020

Abstract

Abstract nnnnn nnnnnn

```
# Explore relationship between income and every major perecntage
education_percentage_only %>%
  gather(-c(median_household_total_income, total_population.x), key = "numberhood_number", value = "valu
  ggplot(aes(x = value, y = as.numeric(median_household_total_income))) +
    facet_wrap(~ numberhood_number, scales = "free") +
    geom_point(shape=20, color="blue", size=1) +
    stat_smooth(method=lm, se=FALSE)
```



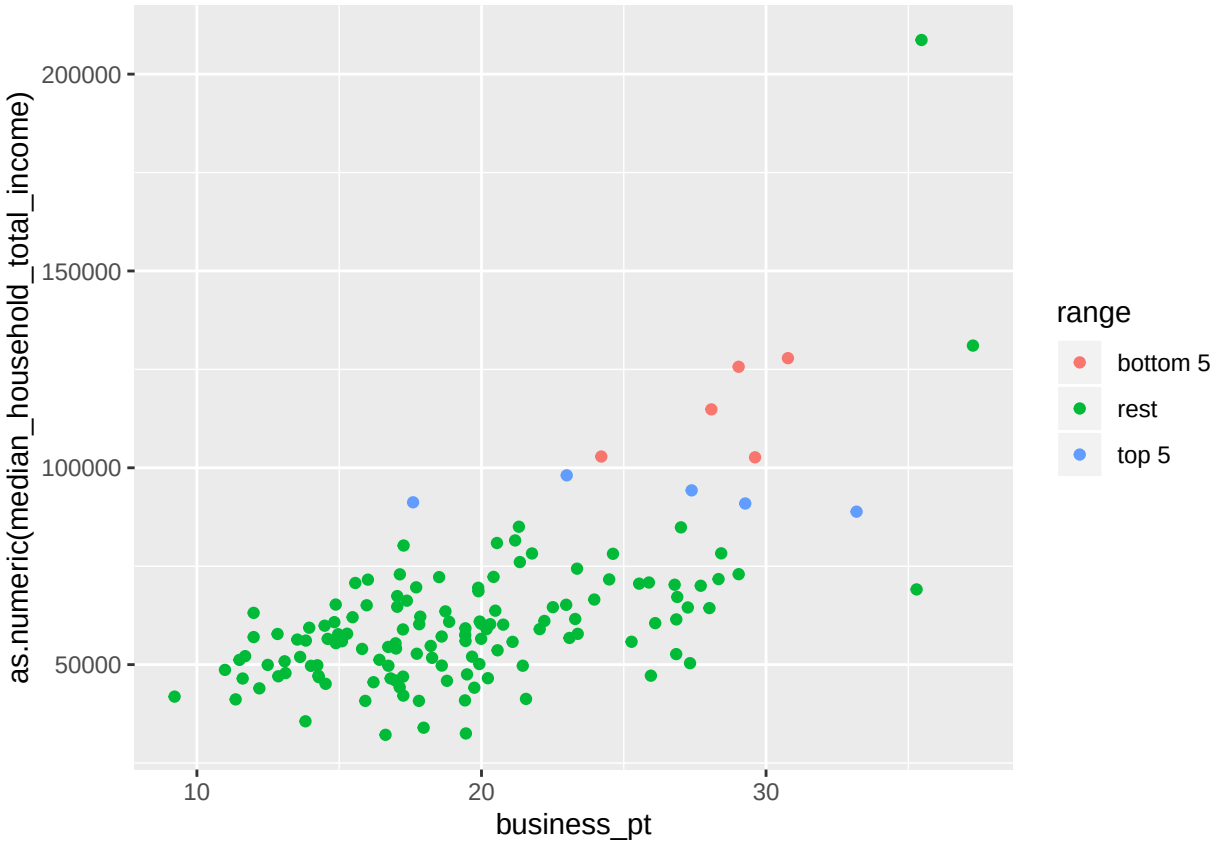

```
# Business major seem to have a stiff line, lets zoom in to it

# Get highest 5 and lowest 5 income values
income <- education_percentage_only$median_household_total_income
h5th <- sort(income)[length(income)-4]
l5th <- sort(income)[5]

# Assign income labels to the neighbourhoods
```

```r
education_percentage_only$range<-ifelse(income >= h5th,"top 5",
        ifelse(income <= l5th,"bottom 5","rest"
        ))

# Plot it
ggplot(education_percentage_only, aes(x=business_pt, y=as.numeric(median_household_total_income), color=
```



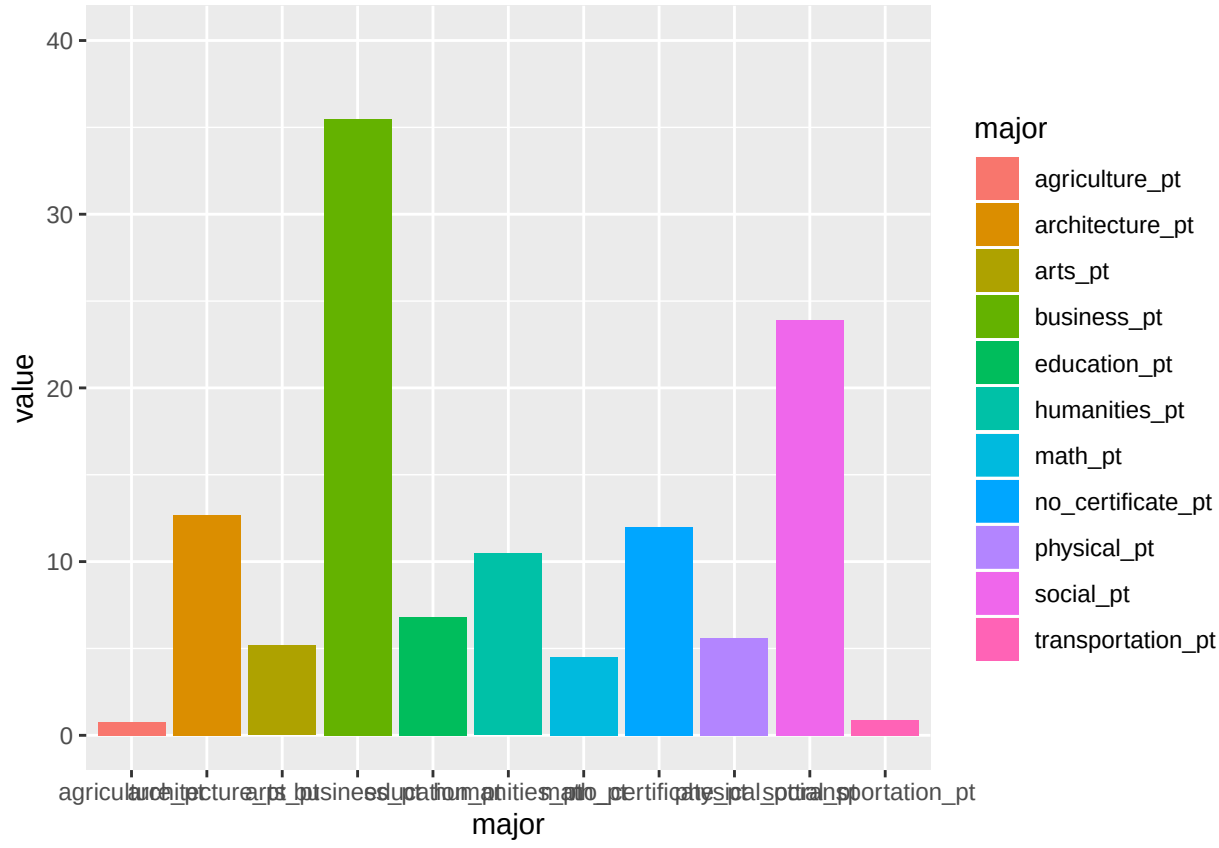

```r
# Linear regression model
linearMod <- lm(business_pt ~ as.numeric(median_household_total_income), data=education_percentage_only)
summary(linearMod)
```

```
##
## Call:
## lm(formula = business_pt ~ as.numeric(median_household_total_income),
##     data = education_percentage_only)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.0574 -3.4944 -0.2842  2.8396 14.4969
##
## Coefficients:
##                                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)                             9.539e+00  1.134e+00   8.408 4.63e-14
## as.numeric(median_household_total_income) 1.628e-04  1.712e-05   9.512  < 2e-16
##
```

```
## (Intercept)                                  ***
## as.numeric(median_household_total_income) ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.328 on 138 degrees of freedom
## Multiple R-squared:  0.396,  Adjusted R-squared:  0.3916
## F-statistic: 90.48 on 1 and 138 DF,  p-value: < 2.2e-16
```

```r
# Plot major percentage distribution in highest income neighbourhood
education_df_highest <- education_percentage_only %>%
  filter(median_household_total_income == max(as.numeric(median_household_total_income)) )
data_plot_highest <-
  education_df_highest %>%
  pivot_longer(cols = "education_pt":"no_certificate_pt", names_to = "major")

# Make a bar chart
data_plot_highest %>%
  ggplot(aes(x = major, y = value, fill = major)) +
  geom_col()+
  ylim(0,40)
```




```r
# Plot major percentage distribution in lowest income neighbourhood
education_df_lowest <- education_percentage_only %>%
   filter(median_household_total_income == min(as.numeric(median_household_total_income)) )
education_df_lowest
```

```
##   total_population.x neighbourhood_number median_household_total_income
## 1               6555                   72                         32172
##   education_pt  arts_pt humanities_pt social_pt business_pt physical_pt
## 1     2.898551 5.644546      6.636156  13.72998    16.62853    2.822273
##    math_pt architecture_pt agriculture_pt transportation_pt no_certificate_pt
## 1 4.347826        9.458429      0.8390542          2.822273          37.90999
##   range
## 1  rest
```

```r
data_plot_lowest <-
  education_df_lowest %>%
  pivot_longer(cols = "education_pt":"no_certificate_pt", names_to = "major")

# Make a bar chart
data_plot_lowest %>%
  ggplot(aes(x = major, y = value, fill = major)) +
  geom_col()+
  ylim(0,40)
```