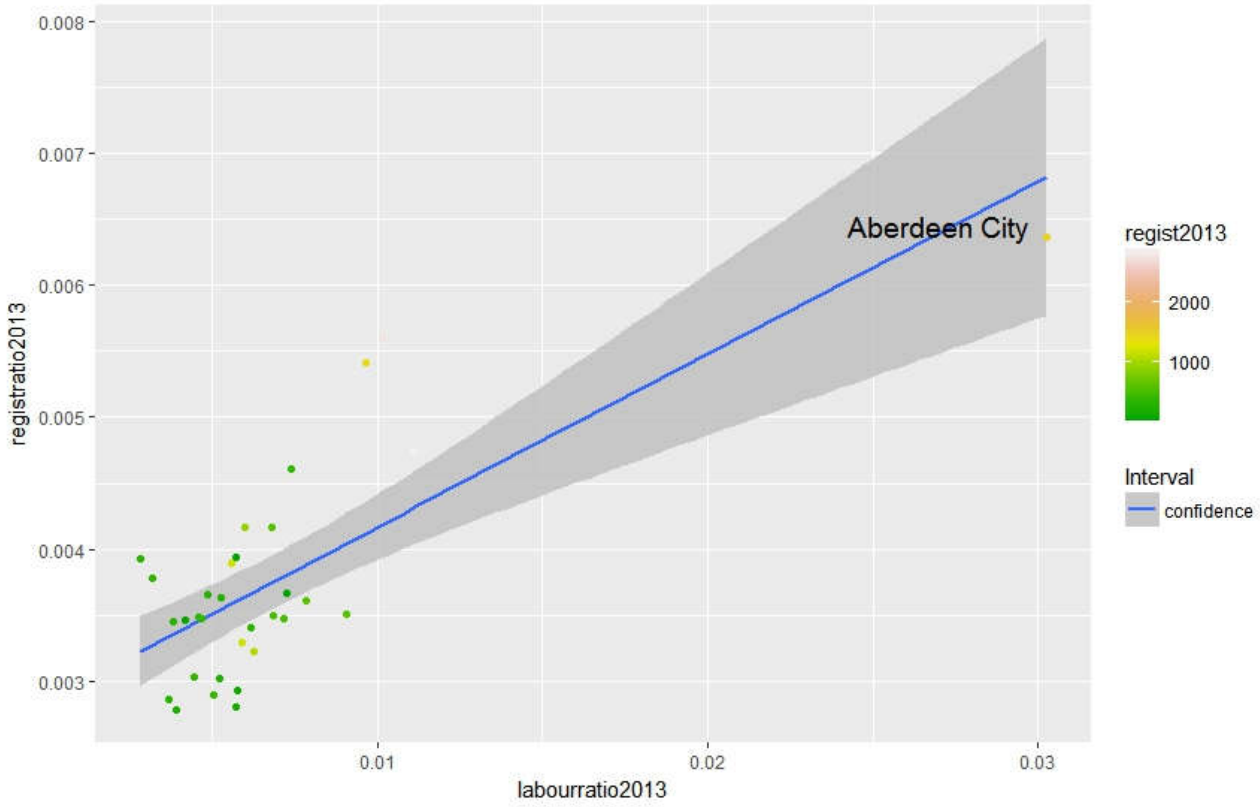
Further questions that come up from our plot.

Can we find the reason we have an extreme value in our plot?

Identifying the extreme value

```
lm_fit = lm(labourratio2013 ~ registratio2013, data = test)
> test = data.frame(test, predict(lm_fit, interval = 'prediction'))
# PLOT WITH REGRESSION LINE, CONFIDENCE INTERVAL AND PREDICTION INTERVAL
p0 <- ggplot(test, aes(x= labourratio2013, y= registratio2013)) +
   geom_smooth(method = 'lm', aes(fill = 'confidence'), alpha = 0.8) +
   geom_point(aes(colour = regist2013)) + scale_colour_gradientn(colours =
terrain.colors(10)) +
   scale_fill_manual('Interval', values = c('grey', 'blue')) +
   geom_text(data=subset(test, labourratio2013 > 0.02), aes(labourratio2013,
registratio2013,label= `Area Name`)  , size = 5, vjust = 0, hjust = 1.1)
p0
```

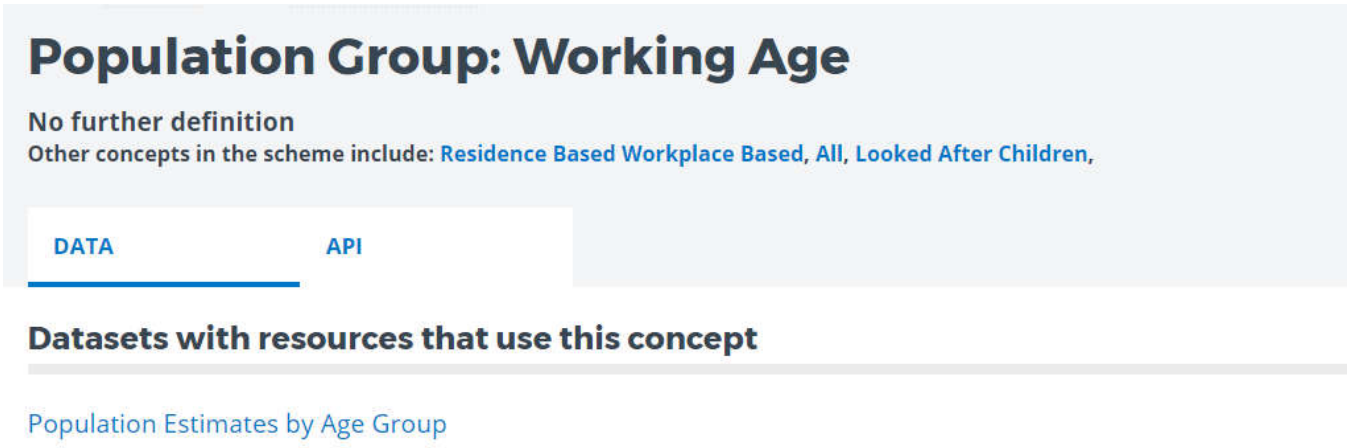


Searching for significance variation in the labor force

First, we are trying to find dataset which could divide our population in different segments

Two similar objects

http://statistics.gov.scot/def/concept/population-group/working-age



http://statistics.gov.scot/def/concept/age/working-age

The first object has more recent data.