The procedure we have followed for a SPARQL query through R in order to get data, to have a regression analysis on them and finally produce a plot with a confidence interval is:

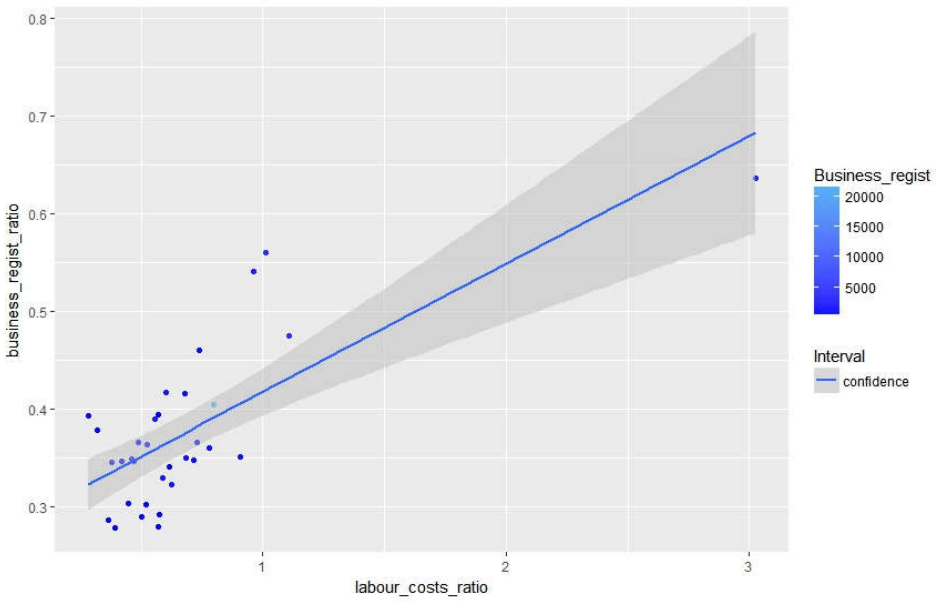
```
#load the needed libraries#
library(RCurl)
library(bitops)
library(XML)
library(SPARQL)
library(ggplot2)

#creation of the query, assigning the results on data frame#
endpoint <- " http://statistics.gov.scot/sparql "
query1 <- " #Our Query with Prefixes (can be found on annex)# "
sp1 <- SPARQL(url=endpoint, query=query1)
test <- data.frame(sp1$results)

#Create ratios#
test["xxxx_ratio"] <- (test$xxxx / test$Pop)*100
[…]
```

	Ref_Area	Labour_costs	Business_regist	Pop	labour_costs_ratio	business_regist_ratio
1	Aberdeen City	6870.0	1445	227070	3.0254987	0.6363676
2	Aberdeenshire	2480.0	1395	257770	0.9620980	0.5411801
3	Angus	533.2	405	116290	0.4585089	0.3482673
4	Argyll and Bute	461.8	320	88050	0.5244747	0.3634299

```
#Create the linear model and its plot with confidence interval#
lm_fit = lm(xxx1_ratio ~ xxx2_ratio, 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= xxx1_ratio, y= xxx2_ratio)) +
    geom_point() +
    geom_point(aes(colour = xxx1)) + scale_colour_gradient(low = "blue") +
    geom_smooth(method = 'lm', aes(fill = 'confidence'), alpha = 0.5) +
    scale_fill_manual('Interval', values = c('grey', 'blue'))
p0
#Optional: summary(lm_fit)#
```



#Our Query with Prefixes#

```
PREFIX qb: <http://purl.org/linked-data/cube#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT ?Ref_Area ?Labour_costs ?Business_regist ?Pop
WHERE {
?a qb:dataSet <http://statistics.gov.scot/data/annual-business-survey>;
<http://statistics.gov.scot/def/dimension/industrySector(sic07)> <http://statistics.gov.scot/def/concept/industry-sector-sic-07/all-sections-a-to-s-as-covered-by-abs> ;
<http://statistics.gov.scot/def/dimension/indicator(abs)> <http://statistics.gov.scot/def/concept/indicator-abs/total-labour-costs>;
        <http://purl.org/linked-data/sdmx/2009/dimension#refPeriod> <http://reference.data.gov.uk/id/year/2013>;
        <http://statistics.gov.scot/def/measure-properties/count> ?Labour_costs;
        <http://purl.org/linked-data/sdmx/2009/dimension#refArea> ?dummy1.
?dummy1 rdfs:label ?Ref_Area.

?b qb:dataSet <http://statistics.gov.scot/data/business-births-deaths-and-survival-rates>;
<http://statistics.gov.scot/def/dimension/indicator(businessBirthsDeathsAndSurvivalRates)>
<http://statistics.gov.scot/def/concept/indicator-business-births-deaths-and-survival-rates/vat/paye-registrations>;
<http://purl.org/linked-data/sdmx/2009/dimension#refPeriod> <http://reference.data.gov.uk/id/year/2013>;
#We link the second dataset to the ref_Area of the first one!!!#
<http://purl.org/linked-data/sdmx/2009/dimension#refArea> ?dummy1;
<http://statistics.gov.scot/def/measure-properties/count> ?Business_regist.

?c qb:dataSet <http://statistics.gov.scot/data/population-estimates-current-geographic-boundaries> ;
<http://statistics.gov.scot/def/dimension/age> <http://statistics.gov.scot/def/concept/age/all> ;
<http://purl.org/linked-data/sdmx/2009/dimension#refPeriod> <http://reference.data.gov.uk/id/year/2013>;
<http://statistics.gov.scot/def/dimension/sex> <http://statistics.gov.scot/def/concept/sex/all>;
<http://purl.org/linked-data/sdmx/2009/dimension#refArea> ?dummy1;
<http://statistics.gov.scot/def/measure-properties/count> ?Pop.

}
ORDER BY ASC(?Ref_Area)
LIMIT 100
```