IGP DATA ANALYSIS

PROJECT TEAM 1

August 22

Data Description: This dataset is part of a larger dataset that has been collected to help to know the impact of Covid 19 Pandemic on the employment profession in the 4 regions of the UK and to also observe the various shifts that occurred in different professions as a result of this, paying particular attention to professions that gained, loss and constant.

```
source("mypackages.R")
```

```
Loading required package: kableExtra

Attaching package: 'flextable'

The following objects are masked from 'package:kableExtra':

    as_image, footnote

Attaching package: 'dplyr'

The following object is masked from 'package:kableExtra':

    group_rows

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union

Attaching package: 'lubridate'

The following objects are masked from 'package:base':

    date, intersect, setdiff, union

Attaching package: 'MASS'

The following object is masked from 'package:dplyr':

    select
```

```
Attaching package: 'psych'

The following objects are masked from 'package:ggplot2':

    %+%, alpha

Attaching package: 'patchwork'

The following object is masked from 'package:MASS':

    area

corrplot 0.92 loaded
```

```
source("helperFunctions.R")

library(tidyr)
```

```
# load dataset
employment_data <- read.csv("merged_data.csv")

#glimpse the loaded csv file to check structure of dataset such as no of rows, no of column and variabl
```

```
library(dplyr)
```

Data Preparation

We are interested in modelling the employment by professions where Country is equal to Scotland.

Rows of data with this property are selected and used to perform the rest of the analysis.

```
#Select rows that satisfies the condition specified in the modelling
employ_scotland<-subset(employment_data,employment_data$Country == "Scotland")
```

```
tinytex::install_tinytex()
```

```
Found 'C:\Users\Haywh\AppData\Roaming\TinyTeX\bin\win32\tlmgr.bat', which indicates a LaTeX distributio

Error in tinytex::install_tinytex(): If you want to force installing TinyTeX anyway, use tinytex::instal
```

```
glimpse(employ_scotland)
```

```
Rows: 18
Columns: 17
$ Date                                                          <int> 2004, ~
$ Country                                                       <chr> "Scotl~
$ X11.Corporate.managers.and.directors                          <int> 132700~
$ X21.Science..research..engineering.and.technology.professionals <int> 100400~
$ X22.Health.professionals                                      <int> 94100,~
$ X23.Teaching.and.educational.professionals                    <int> 108300~
$ X24.Business..media.and.public.service.professionals          <int> 101200~
```

```
$ X33.Protective.service.occupations                                <int> 33500,~
$ X34.Culture..media.and.sports.occupations                         <int> 28500,~
$ X41.Administrative.occupations                                    <int> 239700~
$ X42.Secretarial.and.related.occupations                           <int> 73500,~
$ X51.Skilled.agricultural.and.related.trades                       <int> 34000,~
$ X53.Skilled.construction.and.building.trades                      <int> 93200,~
$ X61.Caring.personal.service.occupations                           <int> 159900~
$ X62.Leisure..travel.and.related.personal.service.occupations      <int> 48100,~
$ X81.Process..plant.and.machine.operatives                         <int> 96900,~
$ X82.Transport.and.mobile.machine.drivers.and.operatives           <int> 91600,~
```

The following exploratory data analysis will be done using descriptive statistics, exploratory graphs and linear correlation.

summary(): To measure central tendency by checking for variation in mean, median, also for outliers. If variation or difference between mean and median is much then it means there's outliers.

Skimr: this is another method used for summarizing data. It gives a broad set of summary statistics with a single function call.

```
summary(employ_scotland)
```

```
     Date        Country           X11.Corporate.managers.and.directors
 Min.   :2004   Length:18          Min.   :132700
 1st Qu.:2008   Class :character   1st Qu.:139375
 Median :2012   Mode  :character   Median :148400
 Mean   :2012                      Mean   :147511
 3rd Qu.:2017                      3rd Qu.:150050
 Max.   :2021                      Max.   :169300
 X21.Science..research..engineering.and.technology.professionals
 Min.   :100400
 1st Qu.:112300
 Median :125200
 Mean   :132767
 3rd Qu.:151425
 Max.   :197700
 X22.Health.professionals X23.Teaching.and.educational.professionals
 Min.   : 94100           Min.   :108300
 1st Qu.:112425           1st Qu.:115325
 Median :120350           Median :116650
 Mean   :119828           Mean   :119744
 3rd Qu.:130150           3rd Qu.:125350
 Max.   :148000           Max.   :133200
 X24.Business..media.and.public.service.professionals
 Min.   : 98900
 1st Qu.:110150
 Median :121450
 Mean   :123606
 3rd Qu.:132500
 Max.   :163800
 X33.Protective.service.occupations X34.Culture..media.and.sports.occupations
 Min.   :26900                      Min.   :28500
 1st Qu.:33750                      1st Qu.:38200
 Median :35950                      Median :42300
```

```
Mean   :35289                      Mean   :44144
3rd Qu.:36750                      3rd Qu.:50475
Max.   :40900                      Max.   :64900
X41.Administrative.occupations X42.Secretarial.and.related.occupations
Min.   :200900                 Min.   :39400
1st Qu.:215525                 1st Qu.:50700
Median :218100                 Median :55950
Mean   :221078                 Mean   :57900
3rd Qu.:226625                 3rd Qu.:69650
Max.   :239700                 Max.   :74000
X51.Skilled.agricultural.and.related.trades
Min.   :29400
1st Qu.:33700
Median :35800
Mean   :34972
3rd Qu.:36200
Max.   :39900
X53.Skilled.construction.and.building.trades
Min.   : 59900
1st Qu.: 87200
Median : 90200
Mean   : 89778
3rd Qu.: 94550
Max.   :108600
X61.Caring.personal.service.occupations
Min.   :159900
1st Qu.:179525
Median :186650
Mean   :185739
3rd Qu.:194200
Max.   :203700
X62.Leisure..travel.and.related.personal.service.occupations
Min.   :48100
1st Qu.:51425
Median :52950
Mean   :54350
3rd Qu.:57750
Max.   :61600
X81.Process..plant.and.machine.operatives
Min.   :58600
1st Qu.:64025
Median :69800
Mean   :72633
3rd Qu.:80925
Max.   :96900
X82.Transport.and.mobile.machine.drivers.and.operatives
Min.   : 77300
1st Qu.: 92925
Median : 99350
Mean   : 97356
3rd Qu.:101350
Max.   :107200
```

```
is.null(employ_scotland)
```

```
[1] FALSE
```

```
skimr::skim(employ_scotland)
```

```
Error in kable_latex(x = structure(c("Name", "Number of rows", "Number of columns", : unused argument (
```

Summary shows there are no outliers as variations between mean and median in all professions are minimal. We checked for null values and got a boolean FALSE.

EXPLORATORY GRAPHS

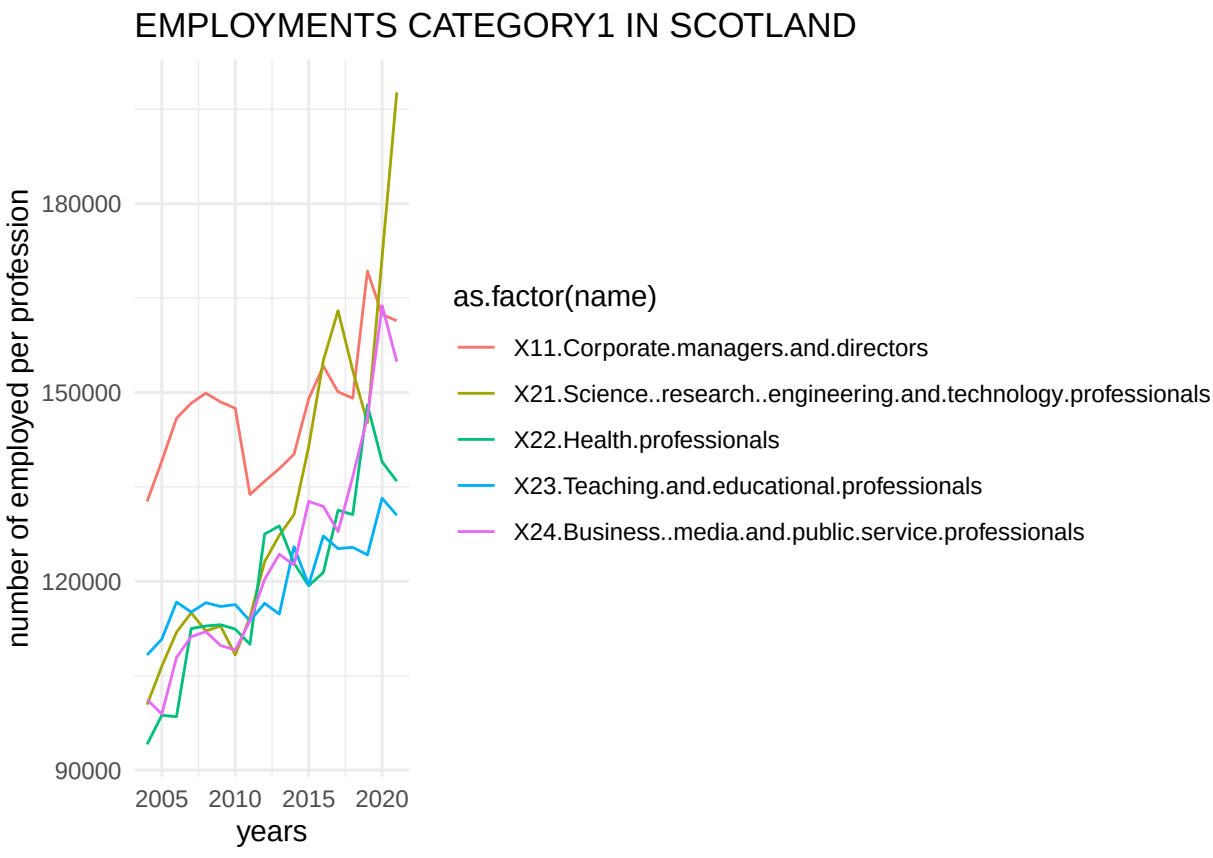
5 professions were selected and grouped into 3 categories and all visually analysed.

```
category1 <- employ_scotland[,c("Date","Country","X11.Corporate.managers.and.directors","X21.Science..re
```

Here we make use of pivot longer which converts wide data into long data using the function pivot longer.Line graphs were plotted showing 5 professions in each categories.

```
#library(tidyverse)
category1_long <- category1%>%
  pivot_longer(
    cols = starts_with("X")
  )
```

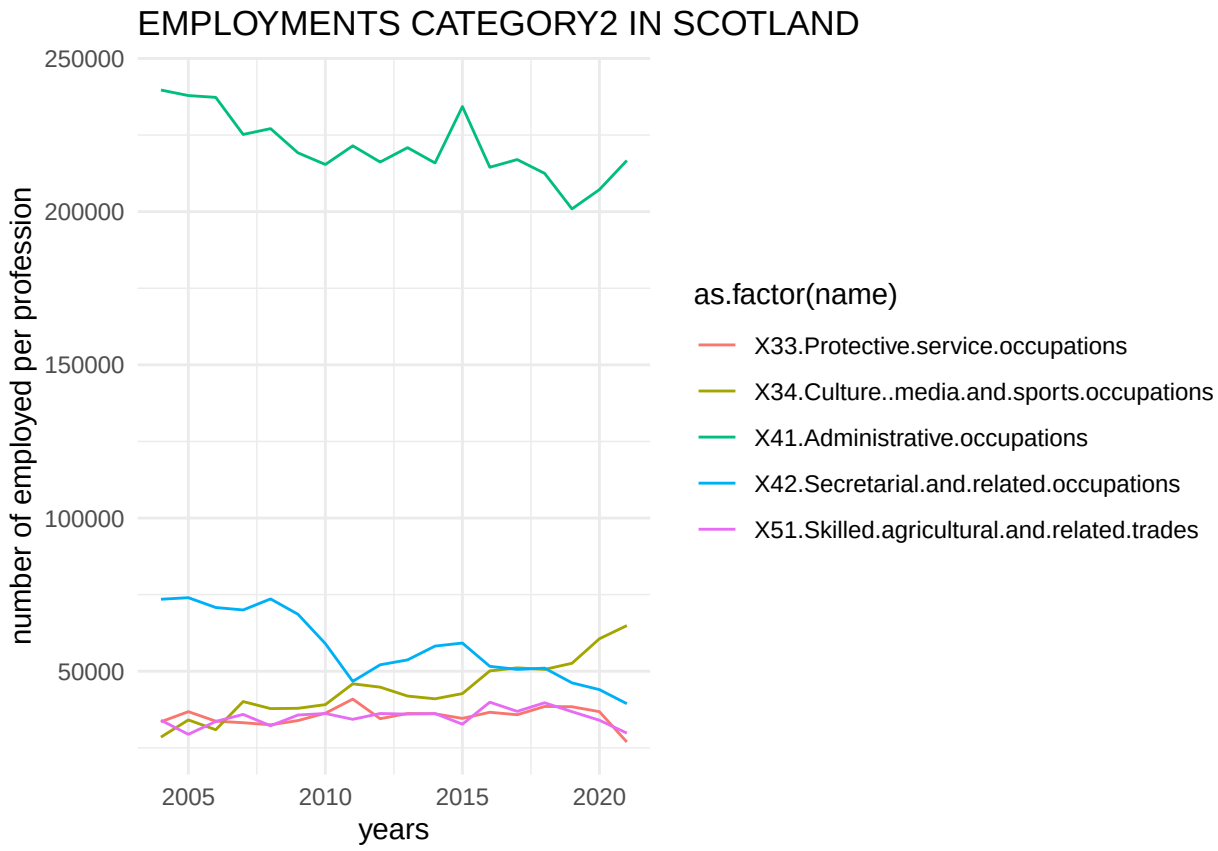
```
ggplot(category1_long,aes(x=Date,y=value,col=as.factor(name)))+geom_line()+xlab("years")+ylab("number o
```



```
category2 <- employ_scotland[,c("Date","Country","X33.Protective.service.occupations", "X34.Culture..me
```

```
category2_long <- category2%>%
  pivot_longer(
    cols = starts_with("X")
  )
```

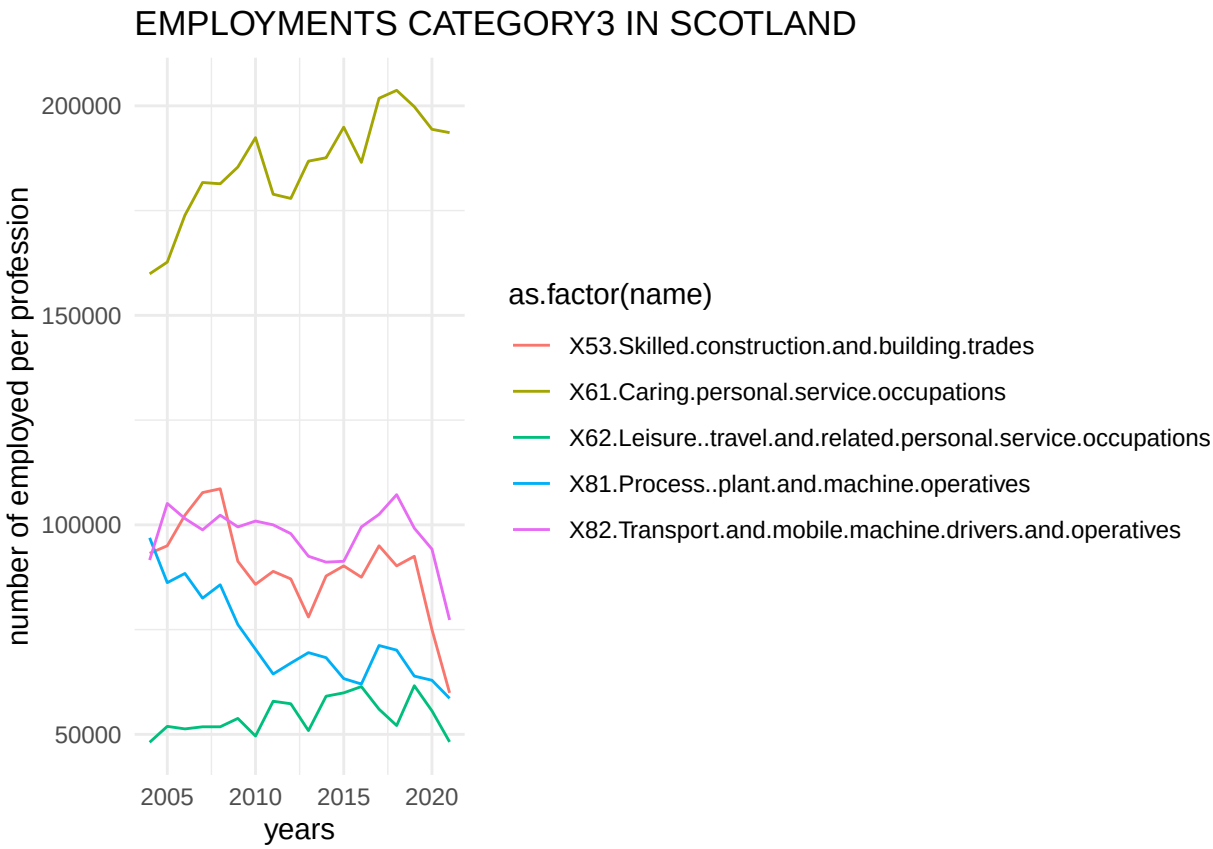
```
ggplot(category2_long,aes(x=Date,y=value,col=as.factor(name)))+geom_line()+xlab("years")+ylab("number o
```




```
category3 <- employ_scotland[,c("Date","Country","X53.Skilled.construction.and.building.trades", "X61.Ca
```

```
category3_long <- category3%>%
  pivot_longer(
    cols = starts_with("X")
  )
```

```
ggplot(category3_long,aes(x=Date,y=value,col=as.factor(name)))+geom_line()+xlab("years")+ylab("number o
```

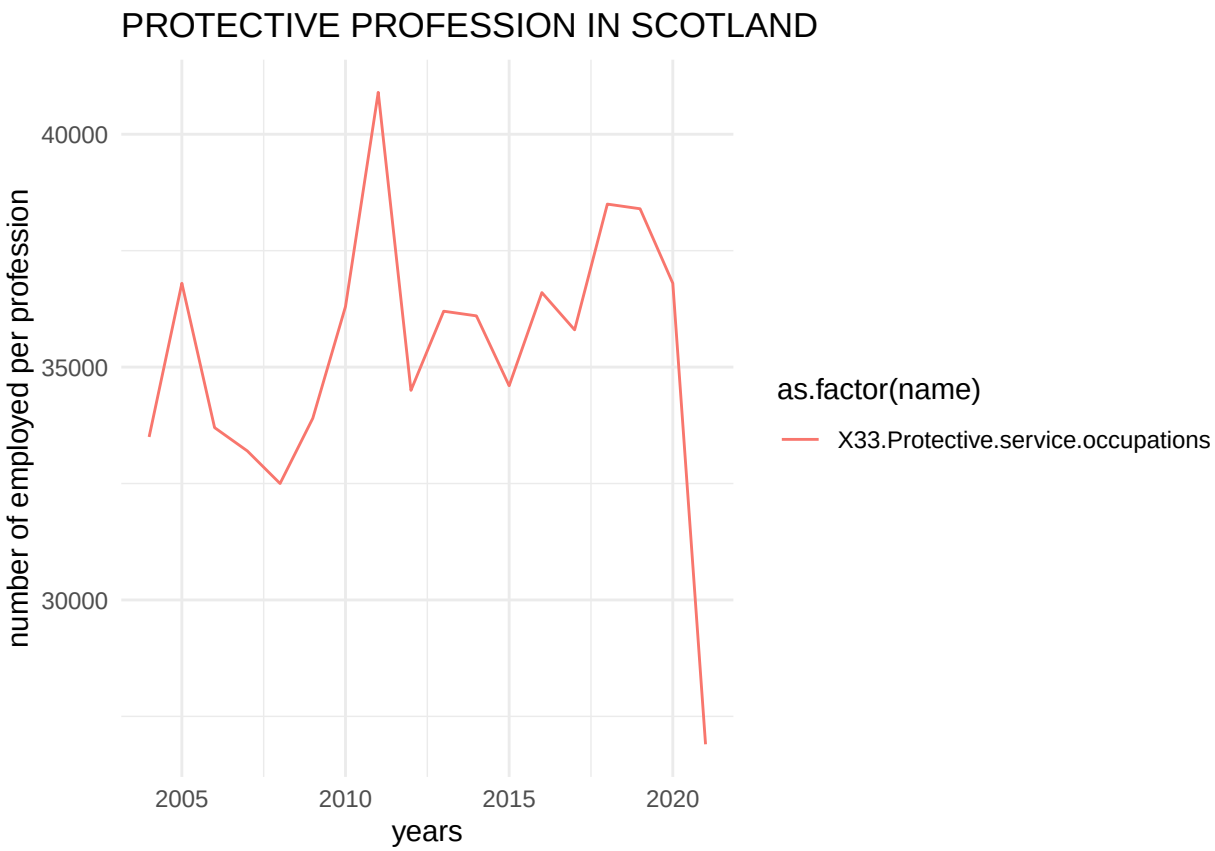



```
protective<- employ_scotland[,c("Date","Country","X33.Protective.service.occupations")]
```

```
protective2 <- protective%>%
  pivot_longer(
    cols = starts_with("X")
  )
```

From the categories plotted, professions with notable changes were plotted individually to further highlight the changes.

```
ggplot(protective2,aes(x=Date,y=value,col=as.factor(name)))+geom_line()+xlab("years")+ylab("number of em
```

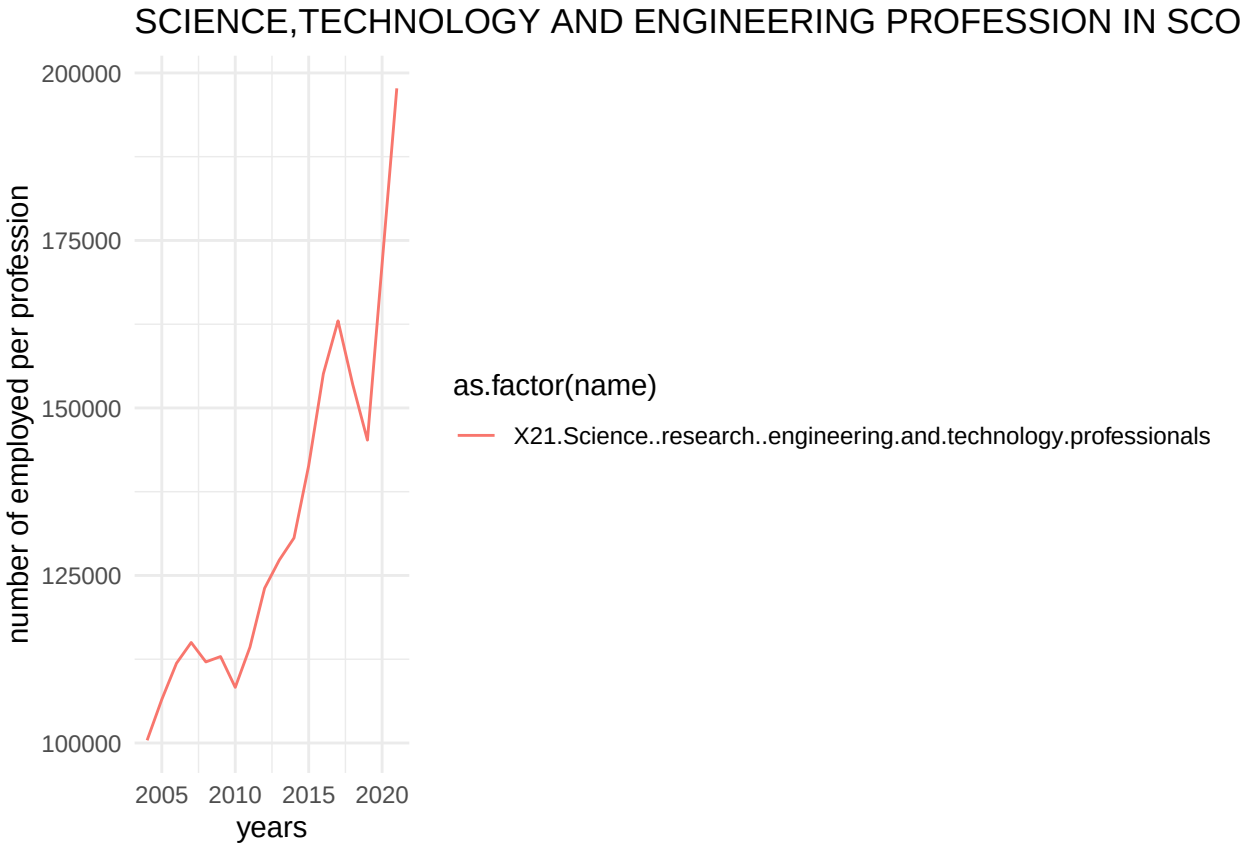


This shows there was a continual decline with effect initializing from 2019 but with significant COVID effect in 2020

```
science<- employ_scotland[,c("Date","Country","X21.Science..research..engineering.and.technology.profess
```

```
science2 <- science%>%
  pivot_longer(
   cols = starts_with("X")
  )
```

```
ggplot(science2,aes(x=Date,y=value,col=as.factor(name)))+geom_line()+xlab("years")+ylab("number of emplo
```

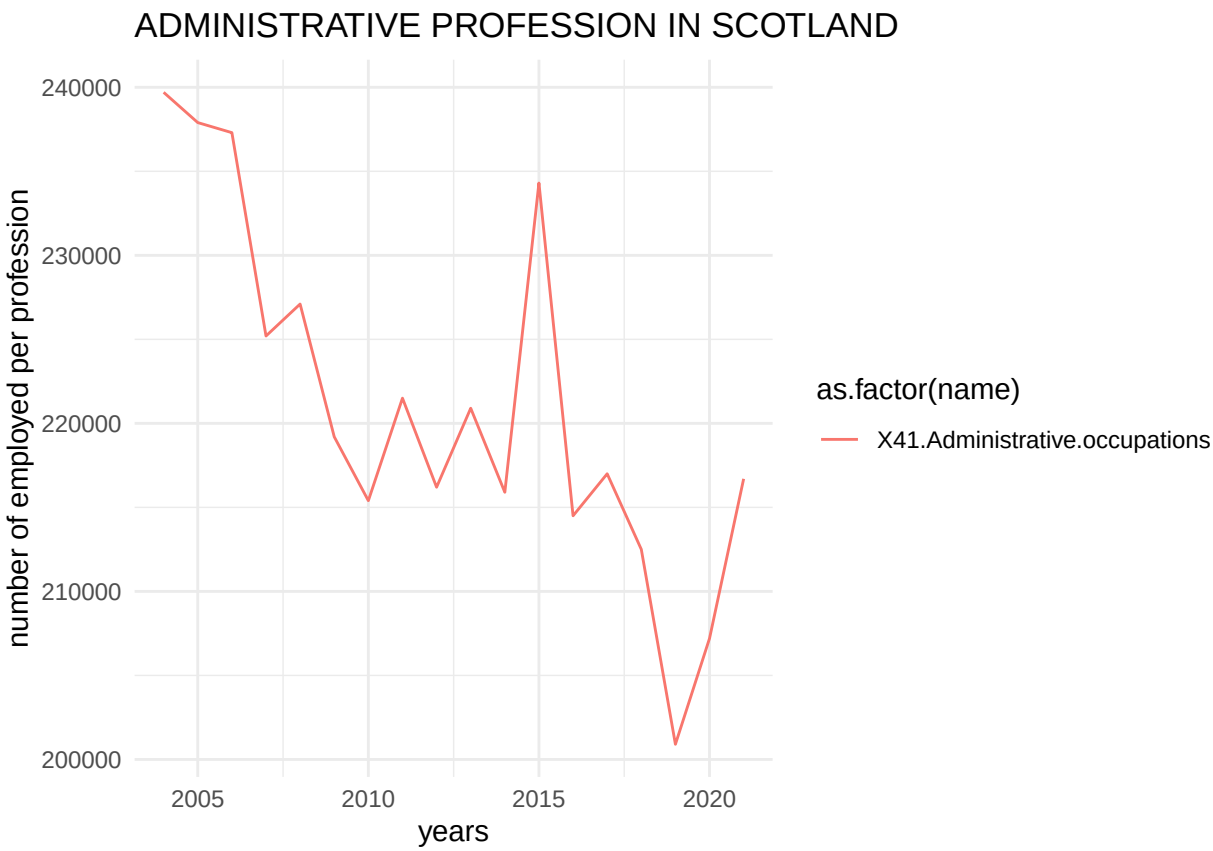


This shows there was a continual increase with effect initializing from 2019 but with significant COVID effect in 2020.

```
admin<- employ_scotland [,c("Date","Country","X41.Administrative.occupations")]
```

```
admin2 <- admin%>%
  pivot_longer(
    cols = starts_with("X")
  )
```

```
ggplot(admin2,aes(x=Date,y=value,col=as.factor(name)))+geom_line()+xlab("years")+ylab("number of employ
```

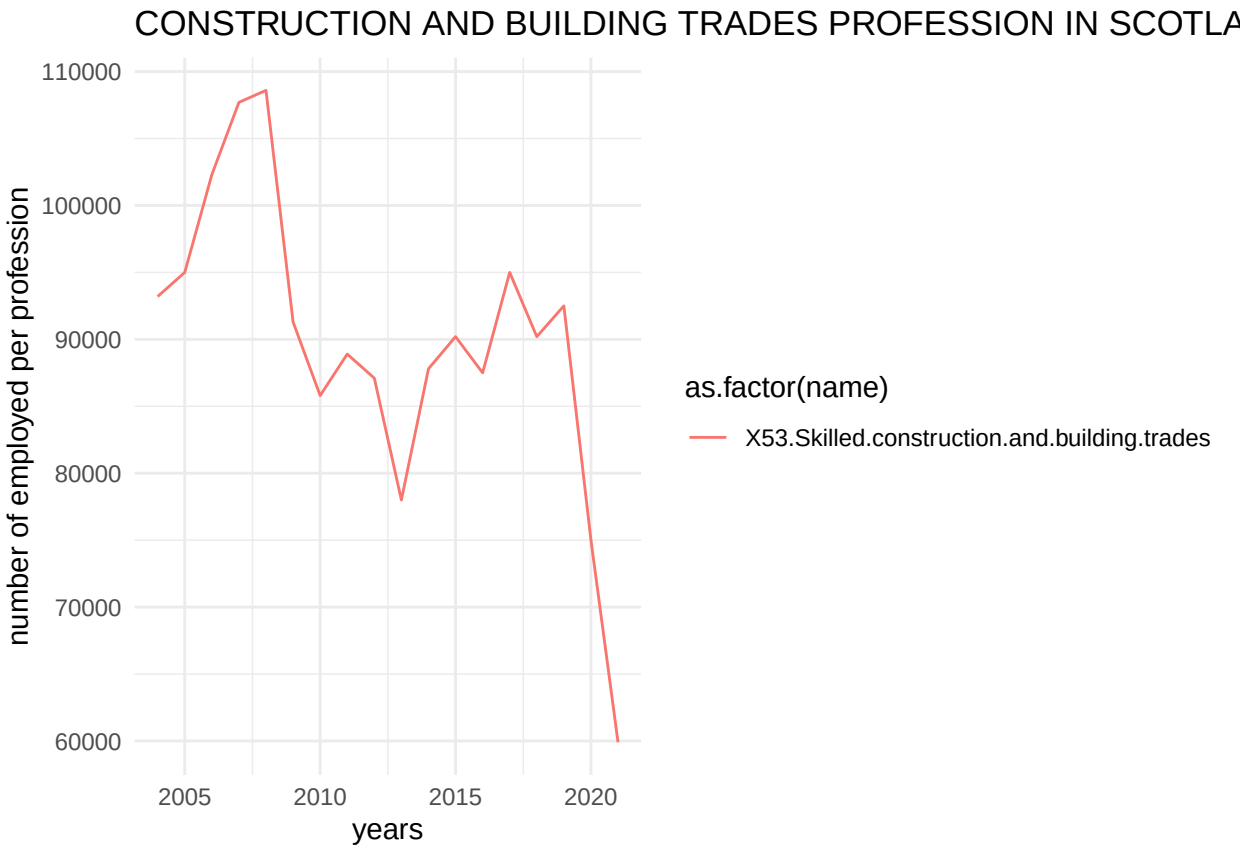


This shows there was a slight increase in 2019 but with significant COVID effect increase in 2020

```
construct<- employ_scotland[,c("Date","Country","X53.Skilled.construction.and.building.trades")]
```

```
construct2 <- construct%>%
  pivot_longer(
    cols = starts_with("X")
  )
```

```
ggplot(construct2,aes(x=Date,y=value,col=as.factor(name)))+geom_line()+xlab("years")+ylab("number of emp
```

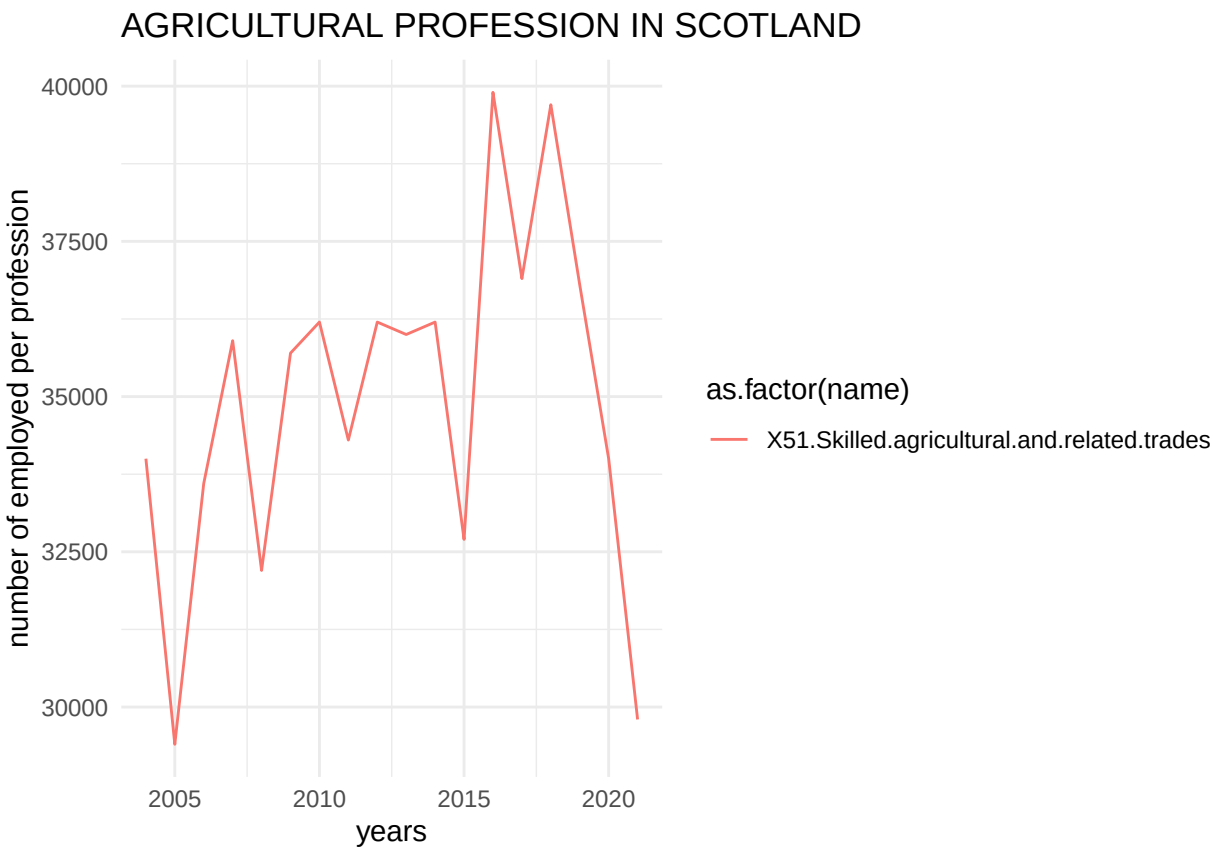


This shows there was a continual decline with effect initializing from 2019 but with significant COVID effect decrease in 2020.

```
agric<- employ_scotland[,c("Date","Country","X51.Skilled.agricultural.and.related.trades")]
```

```
agric2 <- agric%>%
  pivot_longer(
    cols = starts_with("X")
  )
```

```
ggplot(agric2,aes(x=Date,y=value,col=as.factor(name)))+geom_line()+xlab("years")+ylab("number of employ
```



This shows there was a continual decline with effect initializing from 2019 but with significant COVID effect decrease in 2020.

```
caring <- employ_scotland[,c("Date","Country","X61.Caring.personal.service.occupations")]
```

```
caring2 <- caring%>%
  pivot_longer(
    cols = starts_with("X")
  )
```

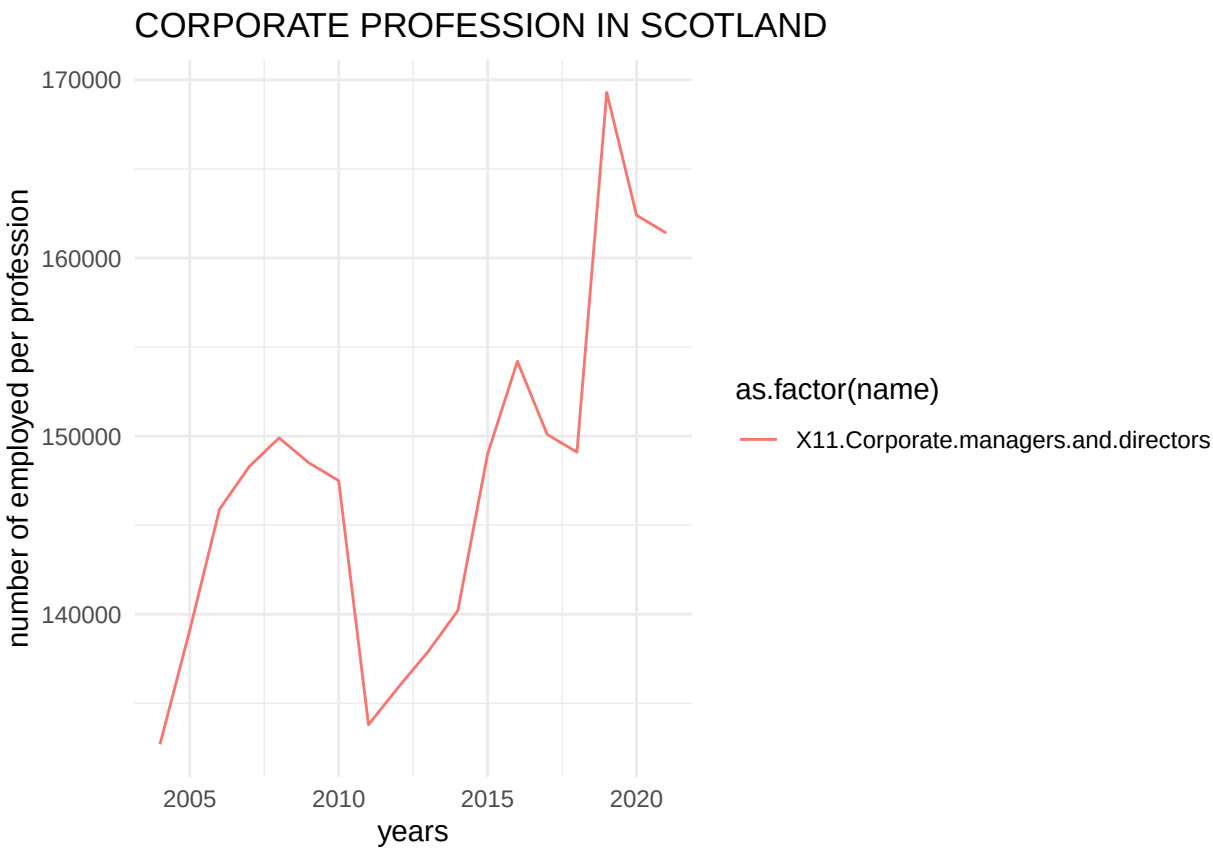
```
ggplot(caring2,aes(x=Date,y=value,col=as.factor(name)))+geom_line()+xlab("years")+ylab("number of emplo
```

This shows there was a decline with effect initializing from 2018 but with a slight change in trend in 2020.

```
corporate <- employ_scotland[,c("Date","Country","X11.Corporate.managers.and.directors")]
```

```
corporate2 <- corporate%>%
  pivot_longer(
    cols = starts_with("X")
  )
```

```
ggplot(corporate2,aes(x=Date,y=value,col=as.factor(name)))+geom_line()+xlab("years")+ylab("number of emp
```

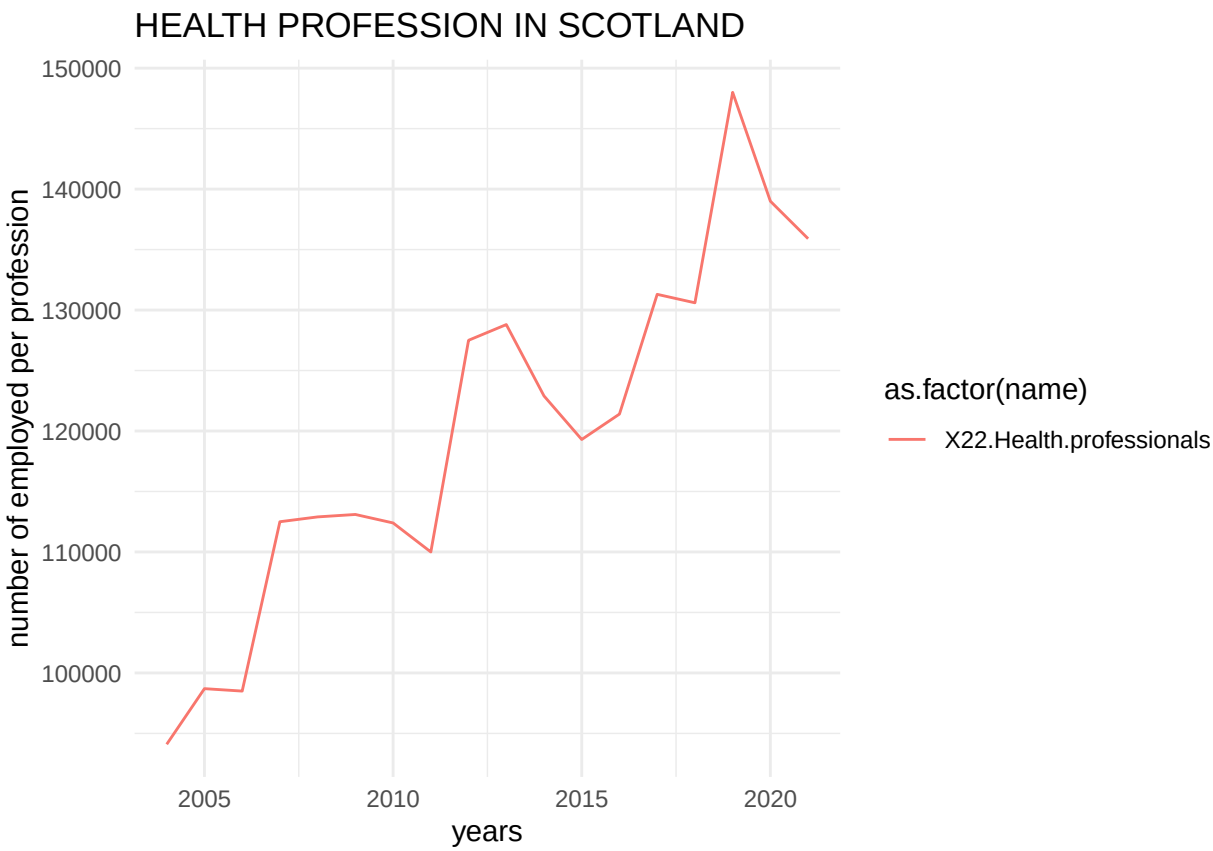


This shows there was a decline with effect initializing from 2018 but with a slight change in trend in 2020

```
health <- employ_scotland[,c("Date","Country","X22.Health.professionals")]
```

```
health2 <- health%>%
  pivot_longer(
    cols = starts_with("X")
  )
```

```
ggplot(health2,aes(x=Date,y=value,col=as.factor(name)))+geom_line()+xlab("years")+ylab("number of emplo
```



This shows there was a decline with effect initializing from 2018 but with almost no significance in 2020.