question2

```r
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   3.5.1     v tibble    3.2.1
v lubridate 1.9.4     v tidyr     1.3.1
v purrr     1.0.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becon
```

```r
library(ggplot2)
library(lme4)  # For linear mixed models
```

```
Loading required package: Matrix

Attaching package: 'Matrix'

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

    expand, pack, unpack
```

```r
facttable<-read_csv("./facttable.csv")
```

```
Rows: 8778 Columns: 24
-- Column specification --------------------------------------------------------
Delimiter: ","
chr  (2): Country Code, Indicator
```

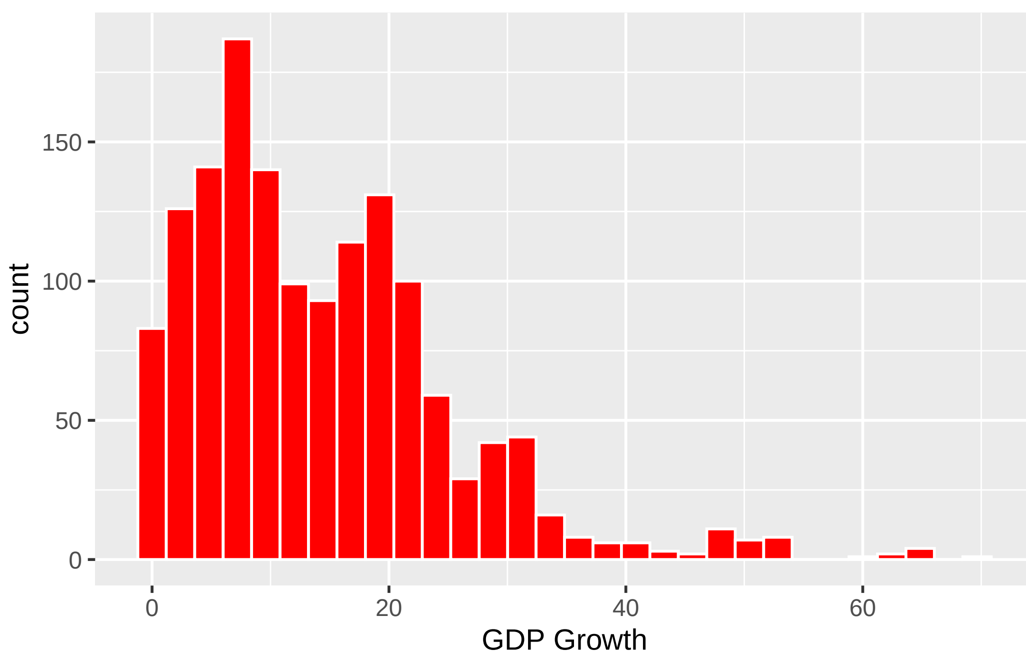
```
dbl (22): 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, ...

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
facttable_wide <- facttable %>%
  pivot_longer(cols = 3:24, names_to = "year", values_to = "value") %>%
  pivot_wider(names_from = Indicator, values_from = value)
```

```r
# Select relevant columns and remove missing values
data <- facttable_wide %>%
  select("Country Code", year, GB.XPD.RSDV.GD.ZS, TX.VAL.TECH.MF.ZS) %>%
  na.omit()
```

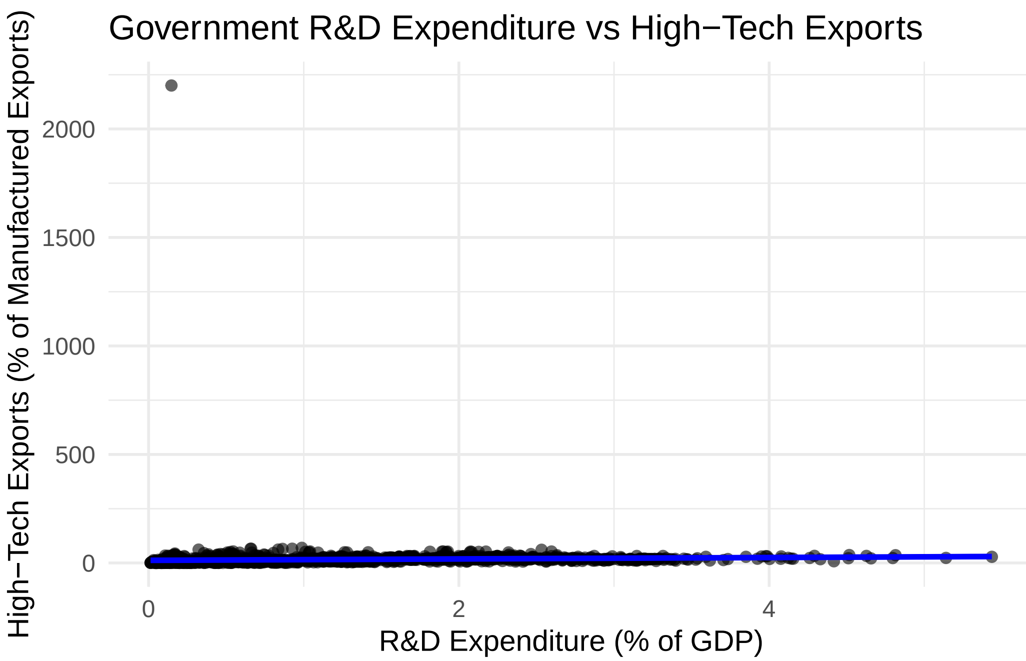
```r
data %>%
  filter(TX.VAL.TECH.MF.ZS < 200) %>%
  drop_na() %>%
  ggplot(aes(x=TX.VAL.TECH.MF.ZS,)) +
  geom_histogram(color="white", fill="red") +
  labs(x = "GDP Growth", colours = "Cylinder")
```

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```r
ggplot(data, aes(x = GB.XPD.RSDV.GD.ZS, y = TX.VAL.TECH.MF.ZS)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", color = "blue") +
  labs(title = "Government R&D Expenditure vs High-Tech Exports",
       x = "R&D Expenditure (% of GDP)",
       y = "High-Tech Exports (% of Manufactured Exports)") +
  theme_minimal()
```

```
`geom_smooth()` using formula = 'y ~ x'
```



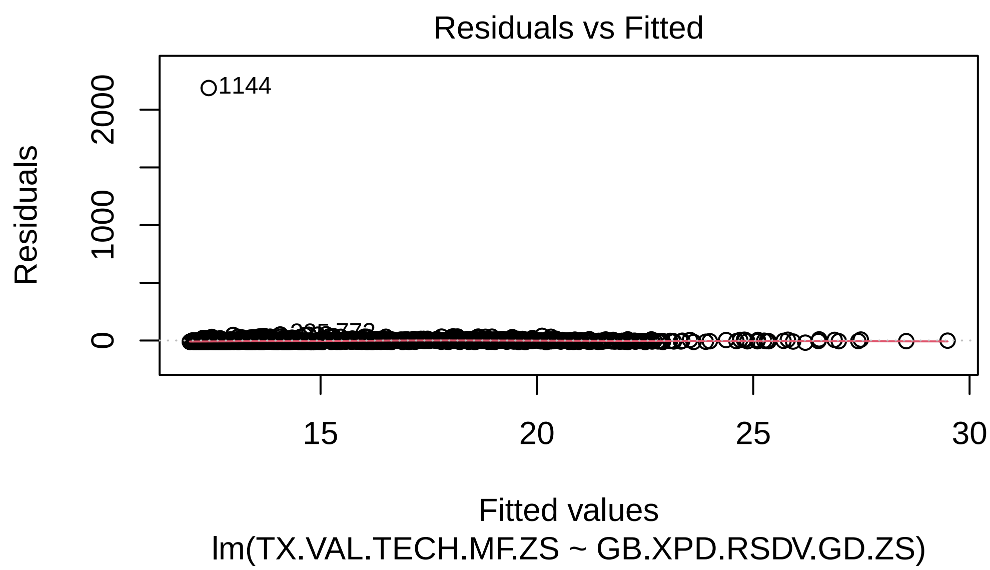
Build a linear model

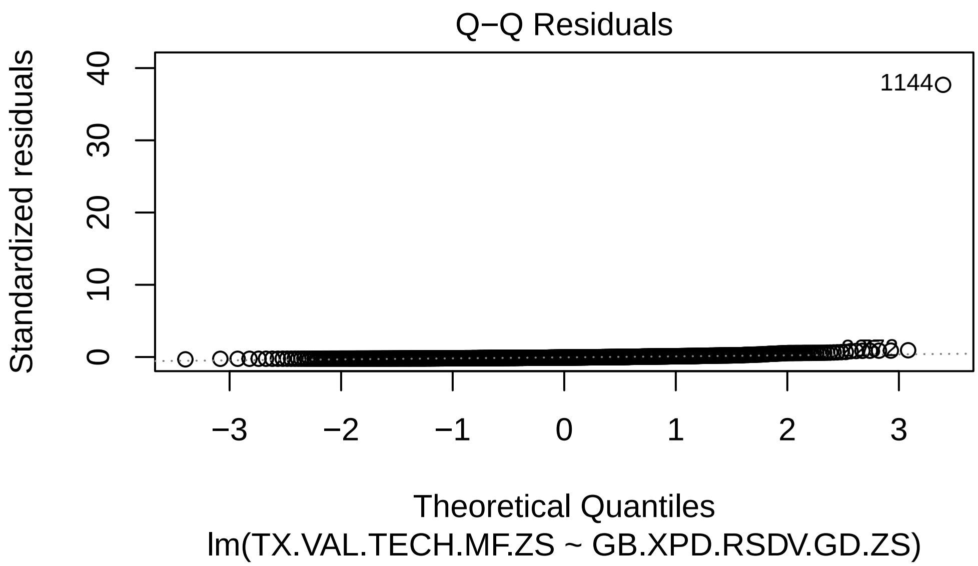
```r
lm_model <- lm(TX.VAL.TECH.MF.ZS ~ GB.XPD.RSDV.GD.ZS, data = data)
summary(lm_model)
```

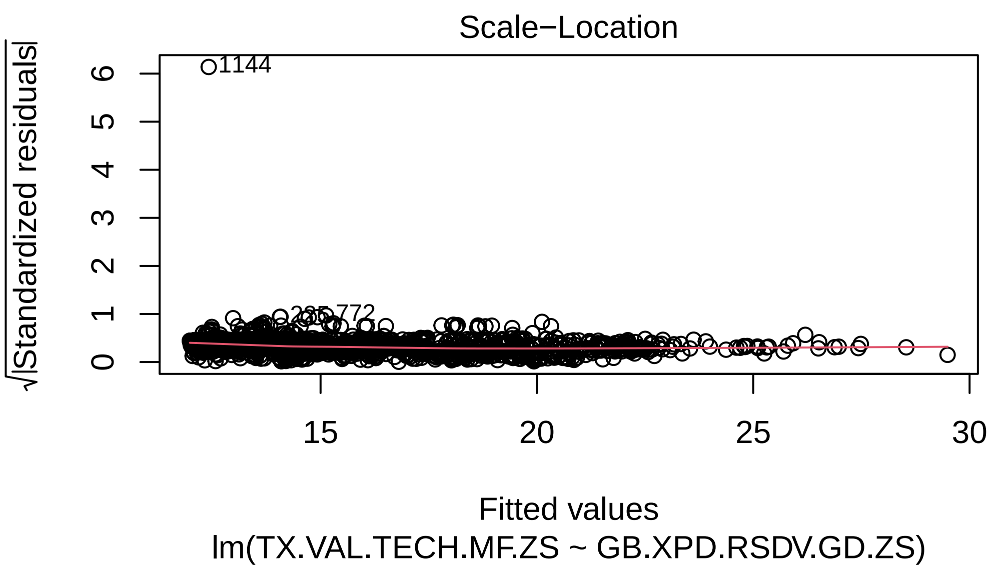
```
Call:
lm(formula = TX.VAL.TECH.MF.ZS ~ GB.XPD.RSDV.GD.ZS, data = data)
```

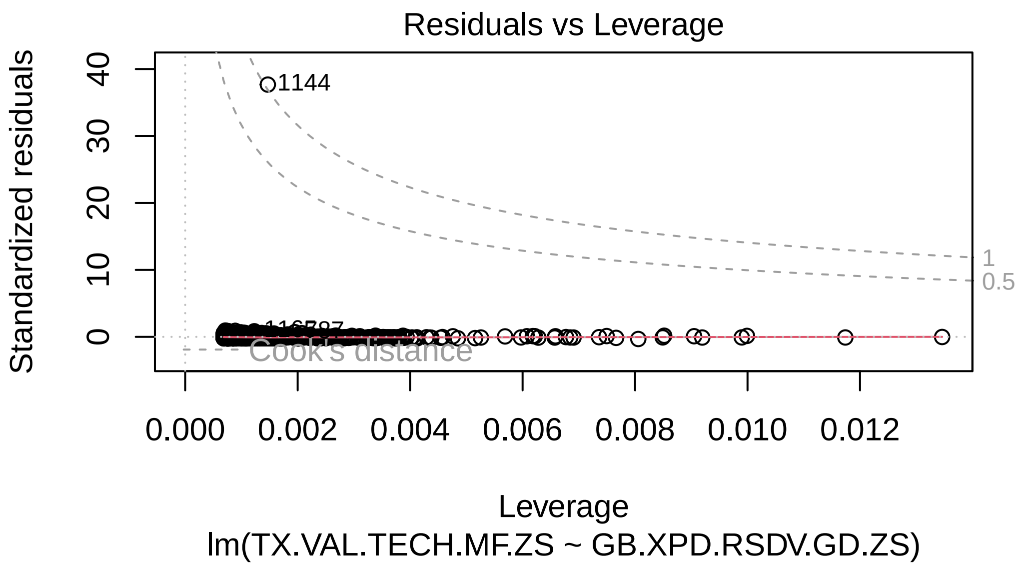
```
Residuals:
    Min      1Q  Median      3Q     Max 
 -18.58   -8.06   -3.98    2.84 2187.83 

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)         11.939      2.399   4.976 7.26e-07 ***
GB.XPD.RSDV.GD.ZS    3.229      1.550   2.083   0.0374 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 58.1 on 1462 degrees of freedom
Multiple R-squared:  0.00296,   Adjusted R-squared:  0.002278 
F-statistic:  4.34 on 1 and 1462 DF,  p-value: 0.0374
```

```r
plot(lm_model)
```




Q−Q Residuals
Standardized residuals
40
30
20
10
0
1144
−3
−2
−1
0
1
2
3
Theoretical Quantiles
lm(TX.VAL.TECH.MF.ZS ~ GB.XPD.RSDV.GD.ZS)


Scale−Location
√|Standardized residuals|
6
5
4
3
2
1
0
1144
772
15
20
25
30
Fitted values
lm(TX.VAL.TECH.MF.ZS ~ GB.XPD.RSDV.GD.ZS)




```r
colnames(data)[colnames(data) == "Country Code"] <- "Country"
```

Building an LMM with Country as Random Effects

```r
lmm_model <- glmer(TX.VAL.TECH.MF.ZS ~ GB.XPD.RSDV.GD.ZS + (1 | Country),  data = data)
```

```
Warning in glmer(TX.VAL.TECH.MF.ZS ~ GB.XPD.RSDV.GD.ZS + (1 | Country), :
calling glmer() with family=gaussian (identity link) as a shortcut to lmer() is
deprecated; please call lmer() directly
```

```r
summary(lmm_model)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: TX.VAL.TECH.MF.ZS ~ GB.XPD.RSDV.GD.ZS + (1 | Country)
   Data: data

REML criterion at convergence: 16036.9

Scaled residuals:
```

```
   Min     1Q Median     3Q    Max 
-1.094 -0.107 -0.050  0.029 37.315 

Random effects:
 Groups   Name        Variance Std.Dev.
 Country  (Intercept)  101.6   10.08  
 Residual             3279.0   57.26  
Number of obs: 1464, groups:  Country, 160

Fixed effects:
                  Estimate Std. Error t value
(Intercept)         11.718      2.715   4.317
GB.XPD.RSDV.GD.ZS    3.327      1.782   1.866

Correlation of Fixed Effects:
            (Intr)
GB.XPD.RSDV -0.765
```

Compare the linear mode with the mixed model to see if indeed there is a random effect.

```r
AIC(lm_model, lmm_model)
```

```
          df      AIC
lm_model   3 16052.90
lmm_model  4 16044.85
```

The comparison of Akaike Information Criterion (AIC) values between the linear model (AIC = 16052.90) and the linear mixed model (AIC = 16044.85) indicates that the mixed-effects model provides a better fit for the data. Since a lower AIC value suggests a more optimal model, this result implies that incorporating random effects to account for country-specific variations improves the explanation of the relationship between government R&D expenditure and high-tech exports. By considering country differences, the mixed model captures heterogeneity more effectively, making it a more suitable choice for understanding how government R&D spending influences high-tech exports across different nations.

Linear Mixed Models

Random Intercept Model

```r
# Random Intercept Model
model_random_intercept <- lmer(TX.VAL.TECH.MF.ZS ~ GB.XPD.RSDV.GD.ZS + (1 | Country), data =
summary(model_random_intercept)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: TX.VAL.TECH.MF.ZS ~ GB.XPD.RSDV.GD.ZS + (1 | Country)
   Data: data

REML criterion at convergence: 16036.9

Scaled residuals:
   Min     1Q Median     3Q    Max
-1.094 -0.107 -0.050  0.029 37.315

Random effects:
 Groups   Name        Variance Std.Dev.
 Country  (Intercept)  101.6   10.08
 Residual             3279.0   57.26
Number of obs: 1464, groups:  Country, 160

Fixed effects:
                  Estimate Std. Error t value
(Intercept)         11.718      2.715   4.317
GB.XPD.RSDV.GD.ZS    3.327      1.782   1.866

Correlation of Fixed Effects:
            (Intr)
GB.XPD.RSDV -0.765
```

This model assumes that each country has a different baseline level of high-tech exports, but the effect of R&D expenditure is the same across all countries(same slope).

Random Intercept and Random Slope Model

```r
#Random Intercept and Random Slope Model

model_random_slope <- lmer(TX.VAL.TECH.MF.ZS ~ GB.XPD.RSDV.GD.ZS + (GB.XPD.RSDV.GD.ZS | Coun
```

```
boundary (singular) fit: see help('isSingular')
```

```r
summary(model_random_slope)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: TX.VAL.TECH.MF.ZS ~ GB.XPD.RSDV.GD.ZS + (GB.XPD.RSDV.GD.ZS |
    Country)
   Data: data

REML criterion at convergence: 16011.1

Scaled residuals:
   Min     1Q Median     3Q    Max
-2.204 -0.095 -0.046  0.030 36.773

Random effects:
 Groups   Name              Variance Std.Dev. Corr
 Country  (Intercept)        407.0   20.173
          GB.XPD.RSDV.GD.ZS   91.6    9.571   -1.00
 Residual                   3185.0   56.436
Number of obs: 1464, groups:  Country, 160

Fixed effects:
                  Estimate Std. Error t value
(Intercept)         11.352      3.085   3.680
GB.XPD.RSDV.GD.ZS    3.767      1.890   1.993

Correlation of Fixed Effects:
            (Intr)
GB.XPD.RSDV -0.821
optimizer (nloptwrap) convergence code: 0 (OK)
boundary (singular) fit: see help('isSingular')
```

This model assumes that both:

- Each country has a different baseline level of high-tech exports (random intercept).
- The effect of R&D expenditure on high-tech exports varies by country (random slope).

```r
# Compare the 2 models;
anova(model_random_intercept, model_random_slope)
```

```
refitting model(s) with ML (instead of REML)
```

```
Data: data
Models:
model_random_intercept: TX.VAL.TECH.MF.ZS ~ GB.XPD.RSDV.GD.ZS + (1 | Country)
model_random_slope: TX.VAL.TECH.MF.ZS ~ GB.XPD.RSDV.GD.ZS + (GB.XPD.RSDV.GD.ZS | Country)
                       npar   AIC   BIC  logLik deviance  Chisq Df Pr(>Chisq)
model_random_intercept    4 16051 16072 -8021.4    16043
model_random_slope        6 16029 16061 -8008.6    16017 25.624  2  2.728e-06

model_random_intercept
model_random_slope     ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The comparison between the random intercept and random slope models shows that the random slope model provides a better fit for the data. The lower AIC (16029 vs. 16051) and BIC (16061 vs. 16072), along with a higher log-likelihood (-8008.6 vs. -8021.4), indicate an improvement in model performance when allowing for random slopes. Additionally, the Chi-square test statistic (25.624, df = 2, p-value = 2.728e-06) confirms that including random slopes significantly enhances the model fit. The highly significant p-value (2.728e-06) suggests that the relationship between government R&D expenditure and high-tech exports varies across countries. Some nations experience a strong positive effect, while others may see weaker or differing impacts. Given this variation, the random slope model is preferred as it captures country-specific differences in how R&D investment translates into high-tech exports. This finding highlights the need to investigate which factors contribute to the differences in slopes among countries, as understanding these variations can provide valuable insights into the conditions under which government R&D spending leads to higher high-tech exports.

```
facttable_wide
```

```
# A tibble: 5,852 x 35
   `Country Code` year  IC.BUS.DISC.XQ IC.CRD.INFO.XQ FS.AST.PRVT.GD.ZS
   <chr>          <chr>          <dbl>          <dbl>             <dbl>
 1 AFE            2000           NA                NA              75.0
 2 AFE            2001           NA                NA              77.0
 3 AFE            2002           NA                NA              62.4
 4 AFE            2003           NA                NA              71.3
 5 AFE            2004           NA                NA              80.3
 6 AFE            2005            3.73             NA              85.8
 7 AFE            2006            3.73             NA              95.0
```

```
 8 AFE            2007          3.73          NA            94.0
 9 AFE            2008          3.69          NA            78.9
10 AFE            2009          3.88          NA            79.3
# i 5,842 more rows
# i 30 more variables: EG.USE.ELEC.KH.PC <dbl>, EG.IMP.CONS.ZS <dbl>,
#   GC.XPN.TOTL.GD.ZS <dbl>, IT.NET.BBND.P2 <dbl>, IT.MLT.MAIN.P2 <dbl>,
#   NY.GDP.MKTP.KD.ZG <dbl>, SI.POV.GINI <dbl>, SE.XPD.TOTL.GD.ZS <dbl>,
#   TX.VAL.TECH.MF.ZS <dbl>, FP.CPI.TOTL.ZG <dbl>, FR.INR.LNDP <dbl>,
#   SL.TLF.CACT.FE.ZS <dbl>, SL.TLF.CACT.MA.ZS <dbl>, LP.LPI.OVRL.XQ <dbl>,
#   CM.MKT.LCAP.GD.ZS <dbl>, MS.MIL.XPND.GD.ZS <dbl>, IT.CEL.SETS.P2 <dbl>, ...
```

```
colnames(facttable_wide)[colnames(facttable_wide) == "Country Code"] <- "Country"
```

```
lmm_extended <- lmer(TX.VAL.TECH.MF.ZS ~ GB.XPD.RSDV.GD.ZS * (CM.MKT.LCAP.GD.ZS + EG.USE.ELEC
                     data = facttable_wide, REML = FALSE)
```

```
Warning: Some predictor variables are on very different scales: consider
rescaling
```

```
summary(lmm_extended)
```

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: TX.VAL.TECH.MF.ZS ~ GB.XPD.RSDV.GD.ZS * (CM.MKT.LCAP.GD.ZS +
    EG.USE.ELEC.KH.PC + FS.AST.PRVT.GD.ZS + GC.TAX.TOTL.GD.ZS +
    IT.MLT.MAIN.P2) + (GB.XPD.RSDV.GD.ZS | Country)
   Data: facttable_wide

     AIC      BIC   logLik deviance df.resid
  2565.8   2633.7  -1266.9   2533.8      496

Scaled residuals:
    Min      1Q  Median      3Q     Max
-5.4321 -0.4068 -0.0499  0.3921  6.7325

Random effects:
 Groups   Name              Variance Std.Dev. Corr
 Country  (Intercept)       185.725  13.628
          GB.XPD.RSDV.GD.ZS  33.498   5.788   -0.80
 Residual                     3.425   1.851
Number of obs: 512, groups:  Country, 82
```

```
Fixed effects:
                                      Estimate Std. Error t value
(Intercept)                         -3.8379872  3.3912534  -1.132
GB.XPD.RSDV.GD.ZS                   11.5309722  2.4854616   4.639
CM.MKT.LCAP.GD.ZS                    0.0287323  0.0100421   2.861
EG.USE.ELEC.KH.PC                    0.0004217  0.0004912   0.859
FS.AST.PRVT.GD.ZS                    0.0096777  0.0229131   0.422
GC.TAX.TOTL.GD.ZS                    0.3230567  0.0650935   4.963
IT.MLT.MAIN.P2                       0.2820679  0.0718262   3.927
GB.XPD.RSDV.GD.ZS:CM.MKT.LCAP.GD.ZS -0.0126617  0.0055888  -2.266
GB.XPD.RSDV.GD.ZS:EG.USE.ELEC.KH.PC -0.0003923  0.0002789  -1.407
GB.XPD.RSDV.GD.ZS:FS.AST.PRVT.GD.ZS  0.0121631  0.0152233   0.799
GB.XPD.RSDV.GD.ZS:GC.TAX.TOTL.GD.ZS -0.3693106  0.0819252  -4.508
GB.XPD.RSDV.GD.ZS:IT.MLT.MAIN.P2    -0.0585382  0.0413888  -1.414

Correlation of Fixed Effects:
                    (Intr) GB.XPD.RSDV.GD.ZS CM.MKT EG.USE FS.AST GC.TAX IT.MLT
GB.XPD.RSDV.GD.ZS   -0.797
CM.MKT.LCAP         -0.125  0.046
EG.USE.ELEC         -0.203  0.095             0.015
FS.AST.PRVT         -0.507  0.423             0.036 -0.249
GC.TAX.TOTL         -0.293  0.391            -0.192 -0.108  0.282
IT.MLT.MAIN         -0.405  0.369            -0.022 -0.397  0.131 -0.095
GB.XPD.RSDV.GD.ZS:C  0.099 -0.042            -0.886  0.002 -0.015  0.210 -0.020
GB.XPD.RSDV.GD.ZS:E  0.184 -0.156            -0.035 -0.868  0.260  0.132  0.333
GB.XPD.RSDV.GD.ZS:F  0.473 -0.511             0.013  0.197 -0.887 -0.251 -0.119
GB.XPD.RSDV.GD.ZS:G  0.257 -0.525             0.158  0.068 -0.208 -0.772  0.009
GB.XPD.RSDV.GD.ZS:I  0.278 -0.327             0.003  0.387 -0.054  0.076 -0.854
                    GB.XPD.RSDV.GD.ZS:C GB.XPD.RSDV.GD.ZS:E GB.XPD.RSDV.GD.ZS:F
GB.XPD.RSDV.GD.ZS
CM.MKT.LCAP
EG.USE.ELEC
FS.AST.PRVT
GC.TAX.TOTL
IT.MLT.MAIN
GB.XPD.RSDV.GD.ZS:C
GB.XPD.RSDV.GD.ZS:E -0.005
GB.XPD.RSDV.GD.ZS:F -0.013              -0.245
GB.XPD.RSDV.GD.ZS:G -0.202              -0.129               0.217
GB.XPD.RSDV.GD.ZS:I  0.046              -0.426               0.023
                    GB.XPD.RSDV.GD.ZS:G
GB.XPD.RSDV.GD.ZS
CM.MKT.LCAP
```

```
EG.USE.ELEC
FS.AST.PRVT
GC.TAX.TOTL
IT.MLT.MAIN
GB.XPD.RSDV.GD.ZS:C
GB.XPD.RSDV.GD.ZS:E
GB.XPD.RSDV.GD.ZS:F
GB.XPD.RSDV.GD.ZS:G
GB.XPD.RSDV.GD.ZS:I  0.001
fit warnings:
Some predictor variables are on very different scales: consider rescaling
```

```r
anova(lmm_extended)
```

```
Analysis of Variance Table
                                    npar  Sum Sq Mean Sq F value
GB.XPD.RSDV.GD.ZS                      1  15.498  15.498  4.5251
CM.MKT.LCAP.GD.ZS                      1  12.911  12.911  3.7699
EG.USE.ELEC.KH.PC                      1  12.321  12.321  3.5974
FS.AST.PRVT.GD.ZS                      1   7.736   7.736  2.2588
GC.TAX.TOTL.GD.ZS                      1  51.756  51.756 15.1119
IT.MLT.MAIN.P2                         1 111.891 111.891 32.6703
GB.XPD.RSDV.GD.ZS:CM.MKT.LCAP.GD.ZS    1  36.259  36.259 10.5871
GB.XPD.RSDV.GD.ZS:EG.USE.ELEC.KH.PC    1  21.796  21.796  6.3639
GB.XPD.RSDV.GD.ZS:FS.AST.PRVT.GD.ZS    1  11.740  11.740  3.4278
GB.XPD.RSDV.GD.ZS:GC.TAX.TOTL.GD.ZS    1  69.544  69.544 20.3056
GB.XPD.RSDV.GD.ZS:IT.MLT.MAIN.P2       1   6.851   6.851  2.0004
```

```r
library(car)
```

```
Loading required package: carData


Attaching package: 'car'

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

    recode
```

```
The following object is masked from 'package:purrr':

    some
```

```r
Anova(lmm_extended, type = "III")
```

```
Analysis of Deviance Table (Type III Wald chisquare tests)

Response: TX.VAL.TECH.MF.ZS
                                      Chisq Df Pr(>Chisq)    
(Intercept)                          1.2808  1   0.257747    
GB.XPD.RSDV.GD.ZS                   21.5237  1  3.495e-06 ***
CM.MKT.LCAP.GD.ZS                    8.1864  1   0.004221 ** 
EG.USE.ELEC.KH.PC                    0.7370  1   0.390608    
FS.AST.PRVT.GD.ZS                    0.1784  1   0.672759    
GC.TAX.TOTL.GD.ZS                   24.6310  1  6.942e-07 ***
IT.MLT.MAIN.P2                      15.4220  1  8.598e-05 ***
GB.XPD.RSDV.GD.ZS:CM.MKT.LCAP.GD.ZS  5.1327  1   0.023480 *  
GB.XPD.RSDV.GD.ZS:EG.USE.ELEC.KH.PC  1.9790  1   0.159497    
GB.XPD.RSDV.GD.ZS:FS.AST.PRVT.GD.ZS  0.6384  1   0.424303    
GB.XPD.RSDV.GD.ZS:GC.TAX.TOTL.GD.ZS 20.3212  1  6.547e-06 ***
GB.XPD.RSDV.GD.ZS:IT.MLT.MAIN.P2     2.0004  1   0.157260    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```