SEM Model

09 May 2018

Functions for converting responses to numbers

```
# for (course) Grade
GradePoint <- function(x) {
  case_when(
    x == "A" ~ 4.0,
    x == "A-" ~ 3.7,
    x == "B+" ~ 3.3,
    x == "B" ~ 3.0,
    x == "B-" ~ 2.7,
    x == "C+" ~ 2.3,
    x == "C" ~ 2.0,
    x == "C-" ~ 1.7,
    x == "D+" ~ 1.3,
    x == "D" ~ 1.0,
    x == "F" ~ 0.0,
    TRUE ~ NA_real_
  )
}

# for Q18, Q19
TextRead <- function(x) {
  case_when(
    x == "0 to 1" ~ 0.5,
    x == "2 to 3" ~ 2.5,
    x == "4 to 5" ~ 4.5,
    x == "6 to 7" ~ 6.5,
    x == "8 to 9" ~ 8.5,
    x == "10 to 11" ~ 10.5,
    x == "12" ~ 12,
    TRUE ~ NA_real_
  )
}

# for Q20
RJTime <- function(x) {
  case_when(
    x == "less than 1" ~ 0.5,
    x == "1 to <3" ~ 2,
    x == "3 to <6" ~ 4.5,
    x == "6 to <10" ~ 8,
    x == "10 to <15" ~ 12.5,
    x == "more than 15" ~ 18,
    TRUE ~ NA_real_
  )
}
```

```r
# for Q21
TextEach <- function(x) {
  case_when(
    x == "less than 1" ~ 0.5,
    x == "1 to <2" ~ 1.5,
    x == "2 to <3" ~ 2.5,
    x == "3 to <4" ~ 3.5,
    x == "4 to <5" ~ 4.5,
    x == "more than 5" ~ 5.5,
    TRUE ~ NA_real_
  )
}

# for Q22, Q23
Attend <- function(x) {
  case_when(
    x == "0%" ~ 0,
    x == "1-20%" ~ 0.1,
    x == "21-40%" ~ 0.3,
    x == "41-60%" ~ 0.5,
    x == "61-80%" ~ 0.7,
    x == "81-100%" ~ 0.9,
    TRUE ~ NA_real_
  )
}

EngGood <- function(x) {
  case_when(
    is.na(x) ~ NA_real_,
    x %in% c("DSE", "Others") ~ NA_real_,

    x %in% c(
      "DSE 5 or above",
      "JiangSu Exam >=105",
      "National Exam >=140",
      "Taiwan Exam >=15"
    ) ~ 1,
    TRUE ~ 0
  )
}
```

Read Data

After reading data, extract useful columns and renaming them.

```r
rawdata <- read_csv("RawData/201617MasterData 2018-03-07.csv",
  na = c("BLANK", "NA", "MULT", "#VALUE!", "(4,6)", "", "ERROR #3100", "#N/A")
)

uge <- rawdata %>%
  mutate(
    # background
    EngGood = EngGood(`English Proficiency`),
```

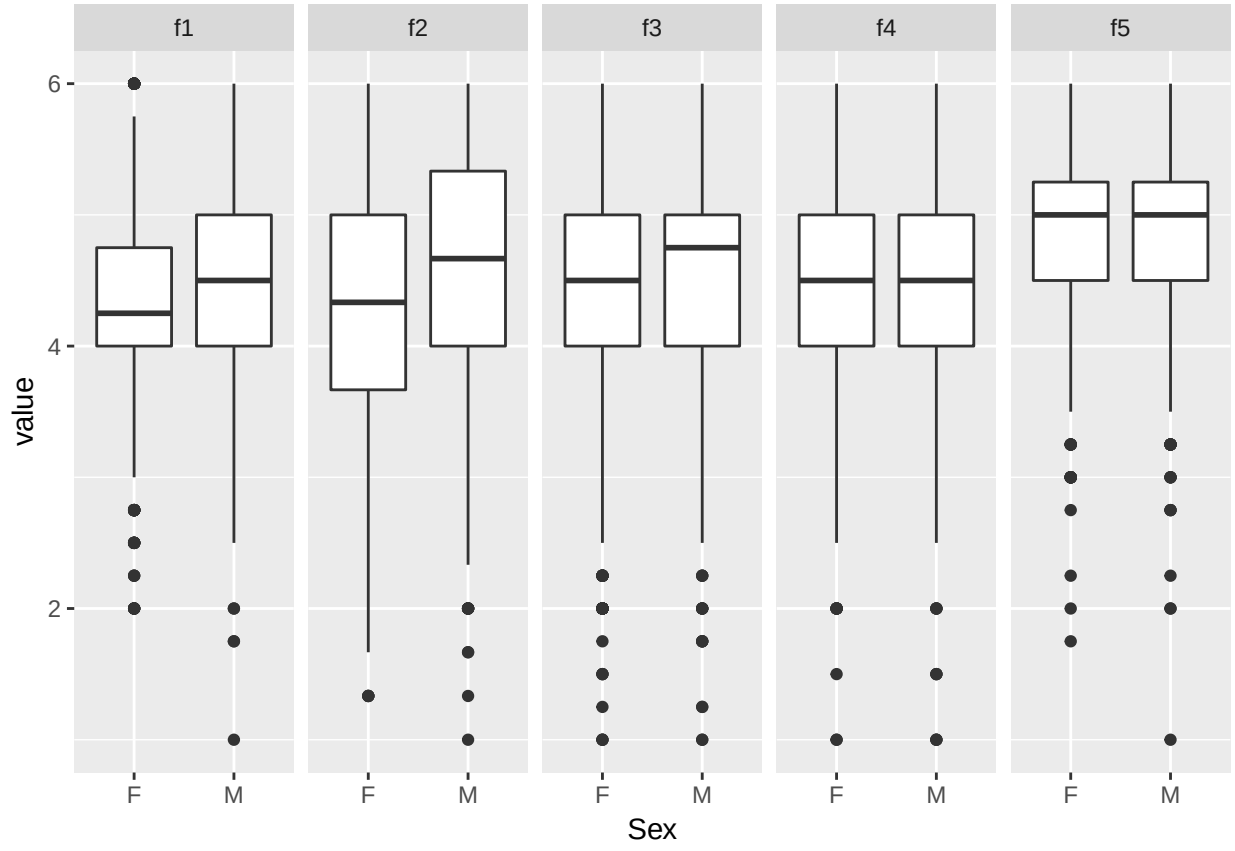
```
    FacSci = Faculty %in% c("BASCI", "ENF", "ENSCF", "SCF", "MED"),
    # Effort
    TextRead = TextRead(`Q18 (Assigned Text Read Completely)`),
    RJTime = RJTime(`Q20 (Time/RJ)`),
    TextTimeEach = TextEach(`Q21 (Time/reading)`),
    Lecture = 13 * Attend(`Q22 (% Lecture)`),
    Tutorial = 13 * Attend(`Q23 (% tutorial participation)`),
    ReadTime = TextRead * TextTimeEach,
    # Grade
    GradePoint = GradePoint(Grade)
  ) %>%
  select(
    Sex,
    SciSubj = `ScienceSubject (Phy,Chem,Bio,InterSci,CombSci)`,
    EngGood,
    Faculty,
    FacSci,
    matches("Before|After"),
    matches("Change"),
    TextRead:ReadTime,
    GradePoint
  )

colnames(uge) <- str_replace(colnames(uge), "(Q\\d*|cGPA)\\s*\\(\\s*Before\\s*\\)", "\\1Entry")
colnames(uge) <- str_replace(colnames(uge), "(Q\\d*|cGPA)\\s*\\(\\s*After\\s*\\)", "\\1Exit")
colnames(uge) <- str_replace(colnames(uge), "(Q\\d*)\\s*\\(\\s*Change\\s*\\)", "\\1Change")
```

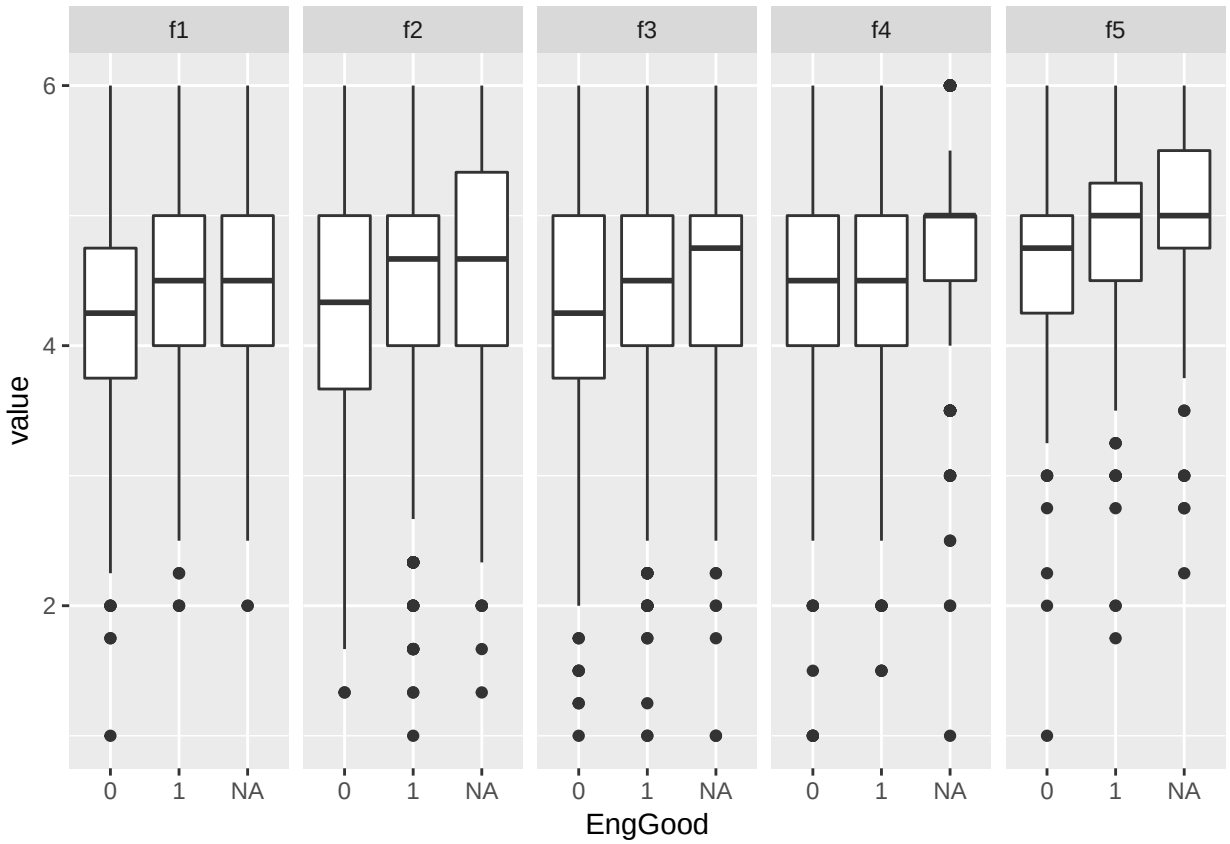
Find backgrounds associated with factors

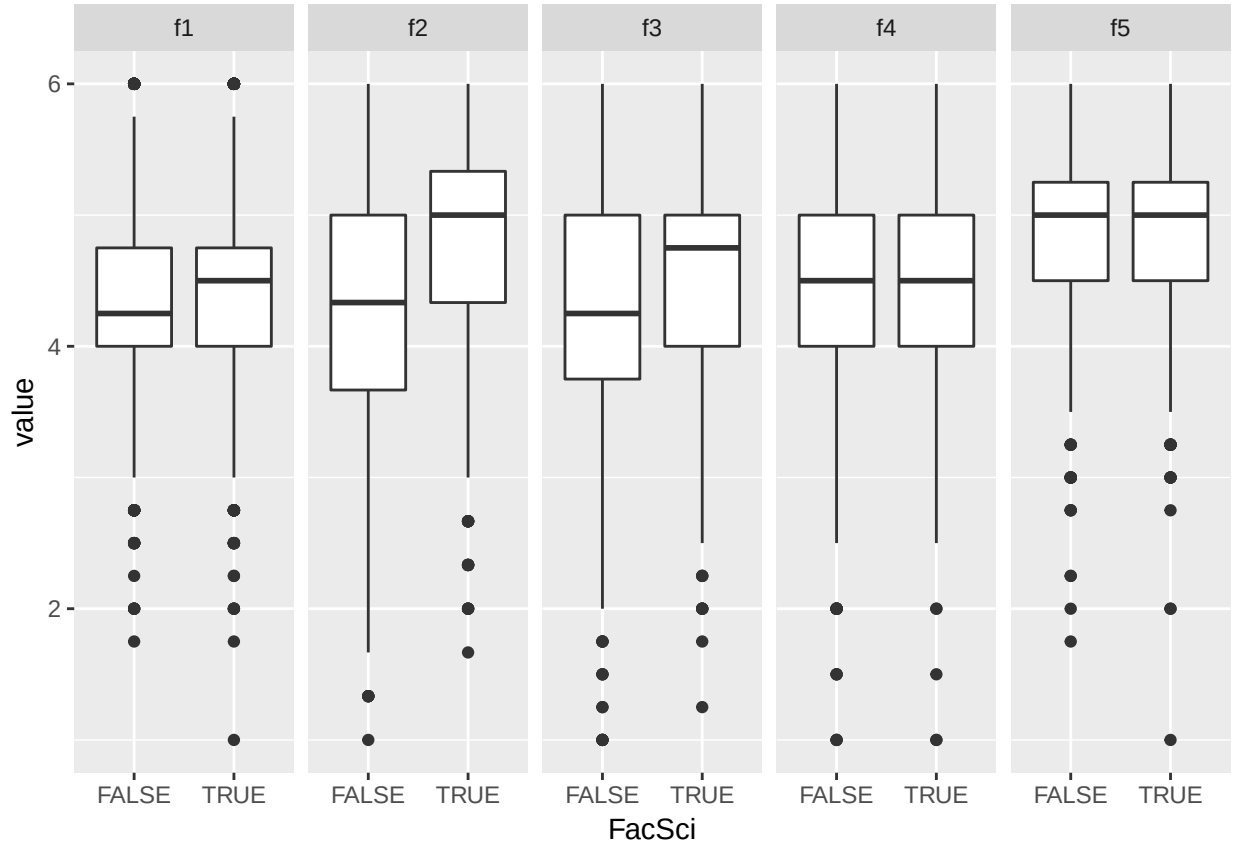
Factors are grouped according to previous years' finding via factor analysis. Backgrounds are chosen based on their correlations with factors before putting them into the model.

```
uge %>% with({
  f1 <- (Q1Exit + Q3Exit + Q4Exit + Q5Exit) / 4
  f2 <- (Q6Exit + Q7Exit + Q8Exit) / 3
  f3 <- (Q9Exit + Q10Exit + Q11Exit + Q12Exit) / 4
  f4 <- (Q13Exit + Q15Exit) / 2
  f5 <- (Q2Exit + Q14Exit + Q16Exit + Q17Exit) / 4

  SexM <- recode(Sex, "F" = 0, "M" = 1)


  # boxplot of factor vs sex
  print(
    tibble(f1, f2, f3, f4, f5, Sex) %>%
      filter(!is.na(Sex)) %>% # 3 NA case only
      gather("factor", "value", -Sex) %>%
      ggplot(aes(Sex, value)) +
      geom_boxplot() +
      facet_grid(. ~ factor)
  )
```

```r
  # boxplot of factor vs English Pro
  print(
    tibble(f1, f2, f3, f4, f5, EngGood = factor(EngGood)) %>%
      gather("factor", "value", -EngGood) %>%
      ggplot(aes(EngGood, value)) +
      geom_boxplot() +
      facet_grid(. ~ factor)
  )

  # boxplot of factor vs Science related faculty
  print(
    tibble(f1, f2, f3, f4, f5, FacSci) %>%
      gather("factor", "value", -FacSci) %>%
      ggplot(aes(FacSci, value)) +
      geom_boxplot() +
      facet_grid(. ~ factor)
  )


  tibble(f1, f2, f3, f4, f5, GradePoint, SciSubj, SexM, EngGood, FacSci) %>%
    cor(use = "pair") %>%
    # only show cor btw factors <-> Grade,backgrounds
    .[1:6, -(1:5)]
})
```




f1
f2
f3
f4
f5
value
6
4
2
0
1
NA
EngGood




```
##             GradePoint    SciSubj        SexM    EngGood       FacSci
## f1           0.2303002 0.11670887  0.09405006 0.20824675  0.016645762
## f2           0.1840714 0.35399737  0.15777333 0.08636422  0.269555867
## f3           0.1804780 0.25642911  0.12690772 0.08193678  0.200176455
## f4           0.1891281 0.06262988  0.05526915 0.09670925  0.008229621
## f5           0.2136377 0.09669791  0.02808321 0.12356411  0.017107120
## GradePoint   1.0000000 0.08532804 -0.06504120 0.26744050 -0.057599146
```

The SEM Model

Specify the model

```
sem.model <- '
  # latent variables
    F1Exit =~ Q1Exit + Q3Exit + Q4Exit + Q5Exit
    F2Exit =~ Q6Exit + Q7Exit + Q8Exit
    F3Exit =~ Q9Exit + Q10Exit + Q11Exit + Q12Exit
    F4Exit =~ Q13Exit + Q15Exit
    F5Exit =~ Q2Exit + Q14Exit + Q16Exit + Q17Exit

    Effort =~ ReadTime + RJTime + Lecture + Tutorial

    Academic =~ GradePoint
```

```
  # regressions (outcome ~ background + effort + other latent var)
    Academic ~ EngGood + Effort + F1Exit + F5Exit
    F1Exit ~ EngGood + Effort
    F2Exit ~ SciSubj + FacSci + Effort
    F3Exit ~ SciSubj + FacSci + Effort + F2Exit
    F4Exit ~ Effort + F5Exit
    F5Exit ~ Effort
'
```

Fit model

The model is fitted and accessed based on CFI and TLI, where a value of 0.9 indicates acceptable fit.

```
sem.fit <- sem(sem.model, data = uge, std.lv = TRUE, missing = "fiml")
(cfi <- fitMeasures(sem.fit, fit.measures = "CFI")) # >.9 okay fit
```

```
##  cfi
## 0.88
```

In this case, the value of CFI is 0.88, which is close to 0.9.

Factor loading Table

```
# nicer table, show composition of factors
parameterEstimates(sem.fit, standardized = TRUE) %>%
  filter(op %in% "=~") %>%
  select("Latent Factor" = lhs, Indicator = rhs, B = est, SE = se,
         Z = z, "p-value" = pvalue, Beta = std.all) %>%
  kable(digits = 3, format = "pandoc", caption = "Factor Loadings")
```

Table 1: Factor Loadings

Latent Factor	Indicator	B	SE	Z	p-value	Beta
F1Exit	Q1Exit	0.286	0.012	23.314	0	0.727
F1Exit	Q3Exit	0.349	0.016	21.655	0	0.803
F1Exit	Q4Exit	0.347	0.016	21.315	0	0.771
F1Exit	Q5Exit	0.372	0.017	21.611	0	0.624
F2Exit	Q6Exit	0.555	0.016	35.514	0	0.809
F2Exit	Q7Exit	0.622	0.018	34.968	0	0.854
F2Exit	Q8Exit	0.482	0.014	34.859	0	0.802
F3Exit	Q9Exit	0.426	0.011	38.480	0	0.856
F3Exit	Q10Exit	0.440	0.011	39.090	0	0.895
F3Exit	Q11Exit	0.409	0.011	38.283	0	0.869
F3Exit	Q12Exit	0.357	0.010	35.656	0	0.782
F4Exit	Q13Exit	0.248	0.016	15.682	0	0.814
F4Exit	Q15Exit	0.242	0.015	15.691	0	0.828
F5Exit	Q2Exit	0.245	0.012	21.026	0	0.657
F5Exit	Q14Exit	0.337	0.016	21.409	0	0.805
F5Exit	Q16Exit	0.325	0.015	21.347	0	0.756

Latent Factor	Indicator	B	SE	Z	p-value	Beta
F5Exit	Q17Exit	0.291	0.015	19.852	0	0.579
Effort	ReadTime	3.036	0.280	10.833	0	0.231
Effort	RJTime	0.369	0.091	4.031	0	0.086
Effort	Lecture	0.636	0.078	8.197	0	0.176
Effort	Tutorial	0.309	0.041	7.501	0	0.160
Academic	GradePoint	0.461	0.018	26.149	0	1.000

Regression Table

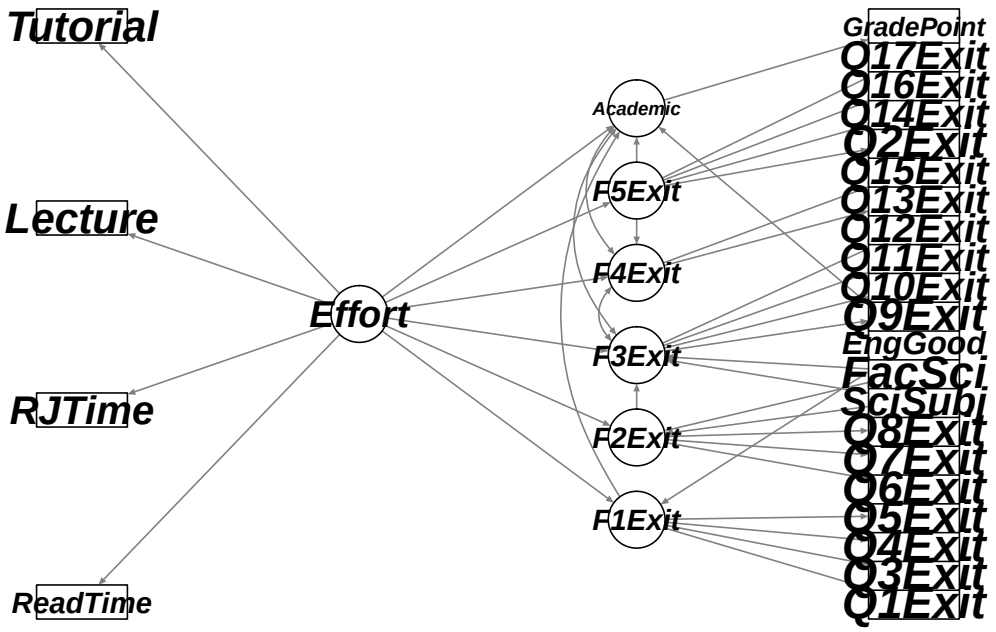
```
# nicer table, show regression of latent variables on measurements/other latent variables
parameterEstimates(sem.fit, standardized = TRUE) %>%
  filter(op %in% "~") %>%
  select("Latent Factor" = lhs, Predictor = rhs, B = est, SE = se,
         Z = z, "p-value" = pvalue, Beta = std.all) %>%
  kable(digits = 3, format = "pandoc", caption = "Regression")
```

Table 2: Regression

Latent Factor	Predictor	B	SE	Z	p-value	Beta
Academic	EngGood	0.676	0.069	9.843	0.000	0.287
Academic	Effort	1.236	0.242	5.107	0.000	1.045
Academic	F1Exit	-0.264	0.064	-4.153	0.000	-0.424
Academic	F5Exit	-0.231	0.066	-3.509	0.000	-0.373
F1Exit	EngGood	0.369	0.069	5.345	0.000	0.097
F1Exit	Effort	1.604	0.091	17.589	0.000	0.845
F2Exit	SciSubj	0.336	0.026	12.844	0.000	0.245
F2Exit	FacSci	0.580	0.057	10.112	0.000	0.189
F2Exit	Effort	1.010	0.047	21.374	0.000	0.660
F3Exit	SciSubj	-0.027	0.028	-0.983	0.325	-0.016
F3Exit	FacSci	0.087	0.060	1.438	0.150	0.022
F3Exit	Effort	0.436	0.054	8.124	0.000	0.227
F3Exit	F2Exit	0.863	0.046	18.931	0.000	0.686
F4Exit	Effort	0.409	0.139	2.939	0.003	0.147
F4Exit	F5Exit	1.169	0.115	10.155	0.000	0.804
F5Exit	Effort	1.628	0.096	17.024	0.000	0.852

Plot Model

```
semPaths(sem.fit, rotation=2,
         residuals = FALSE, exoCov = FALSE, exoVar = FALSE, intercepts = FALSE,
         nCharNodes = 0, font=4, label.cex=2, asize=1.5,
         sizeLat = 5, sizeMan = 8, sizeMan2 =3)
```



For a better image, see the file `semplot.png`.