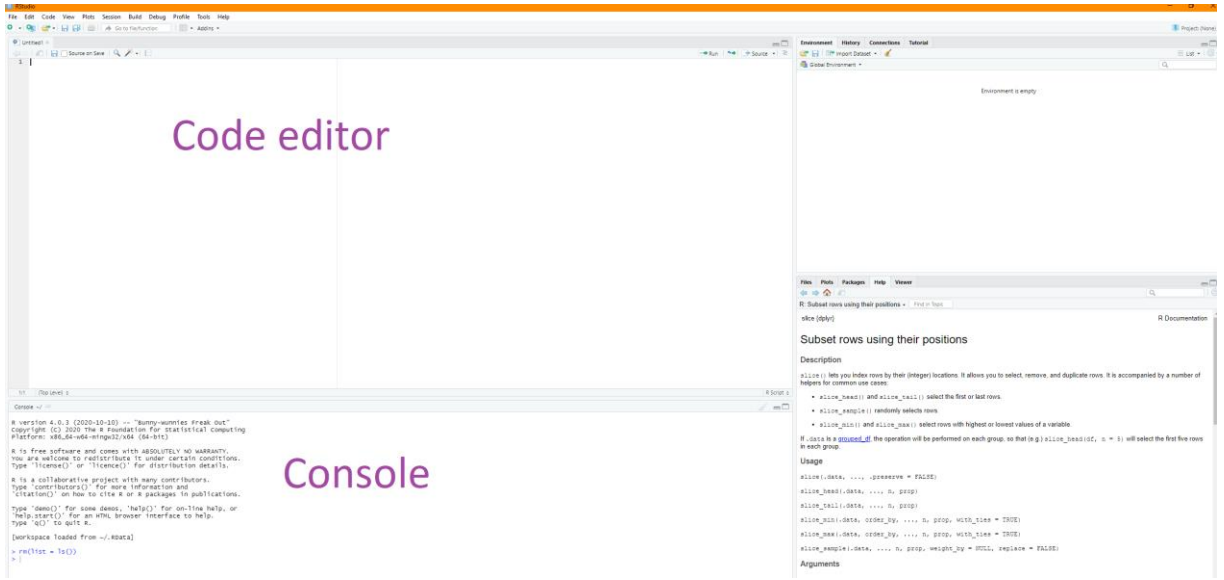
# Intro R

- Open the file day1_intro_multilevel_template.R in Rstudio – we will use this file throughout the day

- Install the packages we need – run the following commands in the console or code editor
  - Install.packages(«lme4»)
  - Install.packages(«tidyverse»)



# Exercises 1

We use the Cassidy dataset with student grade point averages and anxiety levels before taking the test

- GPA – grade point averages
- CTA.tot – anxiety levels

1. Compute the variance of GPA (hint: var())

2. Fit a linear regression model with GPA as dependent variable and anxiety as explanatory variable. Interpret the results (hint: ?lm)

3. What is the estimated residual variance from the regression model? Compare it to the variance you computed before in 1. Why are they different? What is going on? (hint: hist(Cassidy$GPA))

# Exercises 2

We use the Achieve dataset with reading scores from students from different schools

- geread – reading score
- id – student identifier
- school – school identifier

1. How many schools are in the data set? (hint: unique(), length())

2. How many students does each school have on average? What is the standard deviation?

3. Run the code for exercise 2.3. What does it show?

# Exercises 3

lme4 is an R package for fitting multilevel models. The function lmer() can be used to fit linear multilevel models and goes like this:

lmer(y ~ 1 + (1|g), data = mydata)

- y ~ 1 specifies a fixed intercept as in lm()
- (1|g) specifies a random intercept that varies over grouping factor g

We use the Achieve dataset with reading and vocabulary scores from students from different schools

- geread – reading score
- gevocab – vocabulary score
- id – student identifier
- school – school identifier

1. Fit a multilevel model for the reading scores with only an intercept term and random intercepts for schools. Interpret the results (hint: ?lmer)

2. Compute the intraclass correlation for students from the same school. Are differences between schools important for differences between student's reading scores?

3. Include gevocab as an explanatory variable in the model. Interpret the results. Compare the school and residual variance to the "unconditional" model in 1

# Exercises 4

We use the Achieve dataset with reading and vocabulary scores from students from different schools

- geread – reading score
- gevocab– vocabulary score
- id – student identifier
- school – school identifier

1. Expand the last model from the previous exercise to allow for random slopes of gevocab over schools. How did that work? (hint: lmer())

2. Repeat the exercise above but center the vocabulary scores around the grand mean. Interpret the results (hint: x_center = x – mean(x))

3. Perform a likelihood ratio test of whether the random slope is needed in addition to the random intercept (hint: anova(fit1, fit2))

# Exercises 5

We use the Achieve dataset with reading and vocabulary scores from students from different schools

- geread – reading score
- id – student identifier
- school – school identifier
- class – class identifier

1. The class identifier only identifies unique classes within schools. Recode the class identifier so that it takes unique values also across schools (hint: paste0())

2. Fit an «unconditional» 3-level model for the reading scores with random intercepts for classes nested in schools. Interpret the results

3. Compute the intraclass correlation for children from the same schools but different classes. Compute the intraclass correlation for children from the same schools and same classes

# Exercises 6

We use the Achieve dataset with reading and vocabulary scores from students from different schools

- geread – reading score
- gevocab– vocabulary score
- id – student identifier
- school – school identifier
- senroll – number of students in school

1. Include senroll as a covariate for the random school intercepts. Interpret the results

2. Include senroll as a covariate for the random school intercepts and slopes. Interpret the results