how to knitr

me

Abstract

Introduction

This is a minimal example of using **knitr** to produce a *PDF* file from *Markdown*. The original version can be found at the knitr Website.

Related work

Dataset

Exploratory analysis

Add libraries

```
# import libraries
library(psych)
library(corrplot)
library(moments)
```

Now we write some code chunks directly in this markdown file.

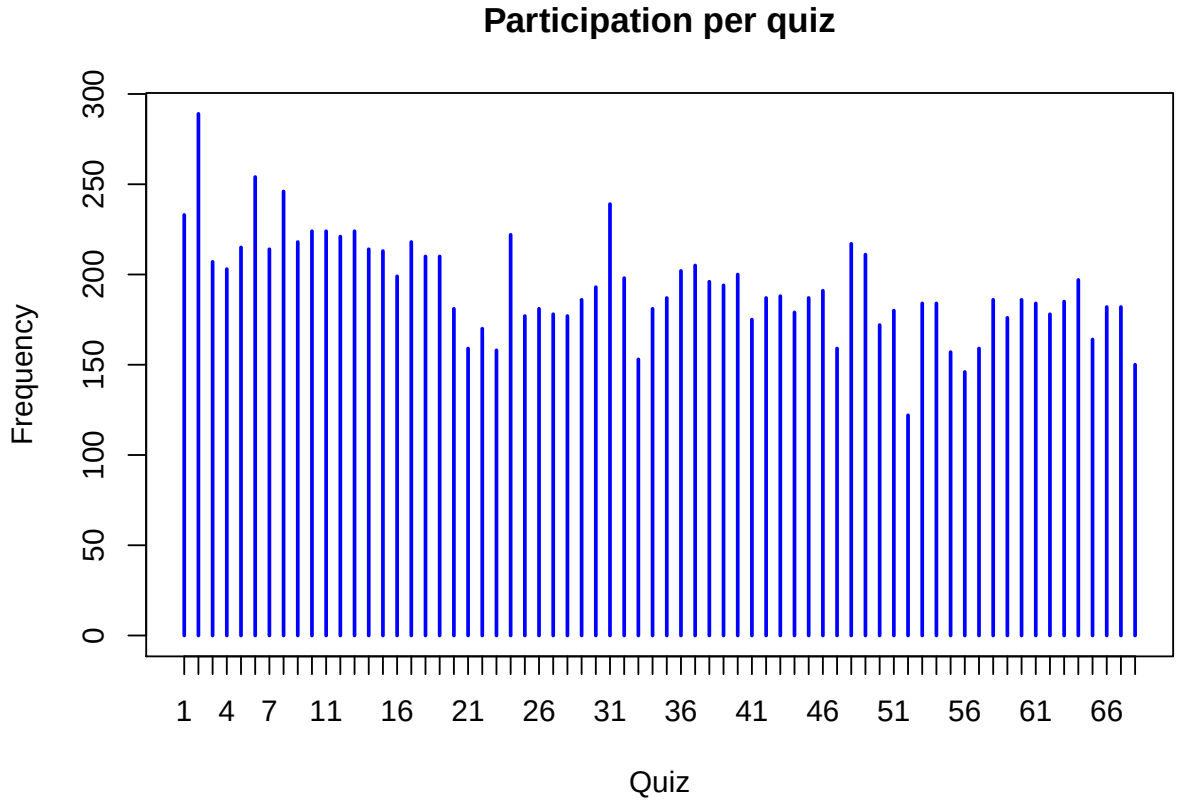
```
# Clean console output and environment
rm(list = ls())
cat("\f")
```

```
# Set local path
path_data <- "C:\\ETH\\StatsLab\\data\\"
# Load files
master_data <- readRDS(paste(path_data, "mastertable.rds", sep = ""))
log_data <- readRDS(paste(path_data, "logDataAnonym.rds", sep = ""))
course_data <- readRDS(paste(path_data, "quizList.rds", sep = ""))
exam_data <- readRDS(paste(path_data, "exam.rds", sep = ""))
```

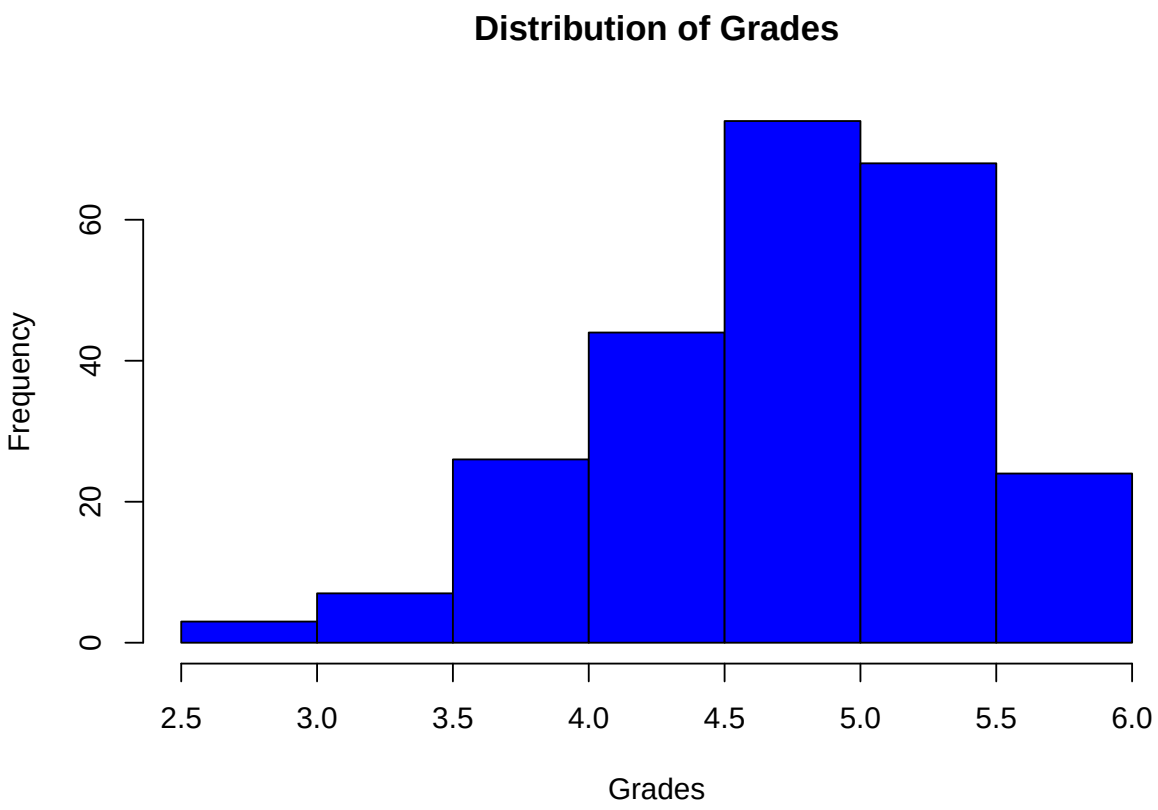
```
# Formatting data - log data
col_log <- c("idVollstName", "idBetroffNutzer", "Zeit", "Ereigniskontext",
    "Komponente", "Ereignisname", "Beschreibung")
log_data <- log_data[, col_log]
log_data$Zeit <- as.POSIXct(log_data$Zeit, format = "%d.%m.%Y %H:%M", tz = "")

# Select ids that have grades
id_students <- unique(c(master_data$id))
log_data <- log_data[log_data$idVollstName %in% id_students, ]

# Difference of students interacted with moodle and those who not
NROW(id_students) - NROW(unique(log_data$idVollstName))
```

```
## [1] 2
```

```
# Remove students who did not interact with moodle and went to the exam
id_students_not_participated <- c(setdiff(id_students, log_data$idVollstName))
exam_data <- exam_data[!(exam_data$id %in% id_students_not_participated),
    ]
master_data <- master_data[!(master_data$id %in% id_students_not_participated),
    ]

# Plot data - participation per quiz
table_quiz <- table(master_data$quiz_nr)
plot(table_quiz, ylab = "Frequency", xlab = "Quiz", col = "blue", main = "Participation per quiz")
```




```
# Plot data - grade distribution
hist(exam_data$Grade, breaks = 10, main = "Distribution of Grades", xlab = "Grades",
    col = "blue")
```




```
# Table summary of grades
grades_dist <- cbind(mean(exam_data$Grade), sd(exam_data$Grade), skewness(exam_data$Grade),
    kurtosis(exam_data$Grade))
points_dist <- cbind(mean(exam_data$Total), sd(exam_data$Total), skewness(exam_data$Total),
    kurtosis(exam_data$Total))
distr <- rbind(grades_dist, points_dist)
colnames(distr) <- c("Mean", "Std. deviation", "Skewness", "Kurtosis")
rownames(distr) <- c("Grades", "Points")
knitr::kable(distr, align = "c")
```

	Mean	Std. deviation	Skewness	Kurtosis
Grades	4.852	0.6588	-0.6355	3.040
Points	16.095	3.1204	-0.8252	3.272

```
# prepare data and remove ids from students and courses
log_data$Beschreibung <- gsub("[[:digit:]]+|The user with id+|id.*|with.*
                              |[[:punct:]]|^\\s+|\\s+$|\n|\\n",
    "", log_data$Beschreibung)
table_analysis <- as.data.frame(table(log_data$idVollstName, log_data$Komponente,
    log_data$Beschreibung))
grades <- as.matrix(exam_data[order(exam_data$id), c("Total")])

# Check length of the matrices
nid <- NROW(unique(table_analysis$Var1))
```

```
ncomp <- NROW(unique(table_analysis$Var2))
nstatus <- NROW(unique(table_analysis$Var3))

# Determine correlations and format row names
comp_names <- paste0(gsub("OU+|\\s+|\\s+$", "", rep(as.character(unique(table_analysis$Var2)))),
    " - ", rep(as.character(unique(table_analysis$Var3)), each = ncomp))
count_test <- matrix(c(table_analysis$Freq), nrow = nid, ncol = ncomp *
    nstatus)
colnames(count_test) <- comp_names
count_select <- count_test[, colSums(count_test) > 5]

# Perform t-test for the correlation
NROW(count_select)
```

```
## [1] 246
```

```
NROW(grades)
```

```
## [1] 246
```

```
test_holm <- corr.test(count_select, grades, adjust = "holm")
```

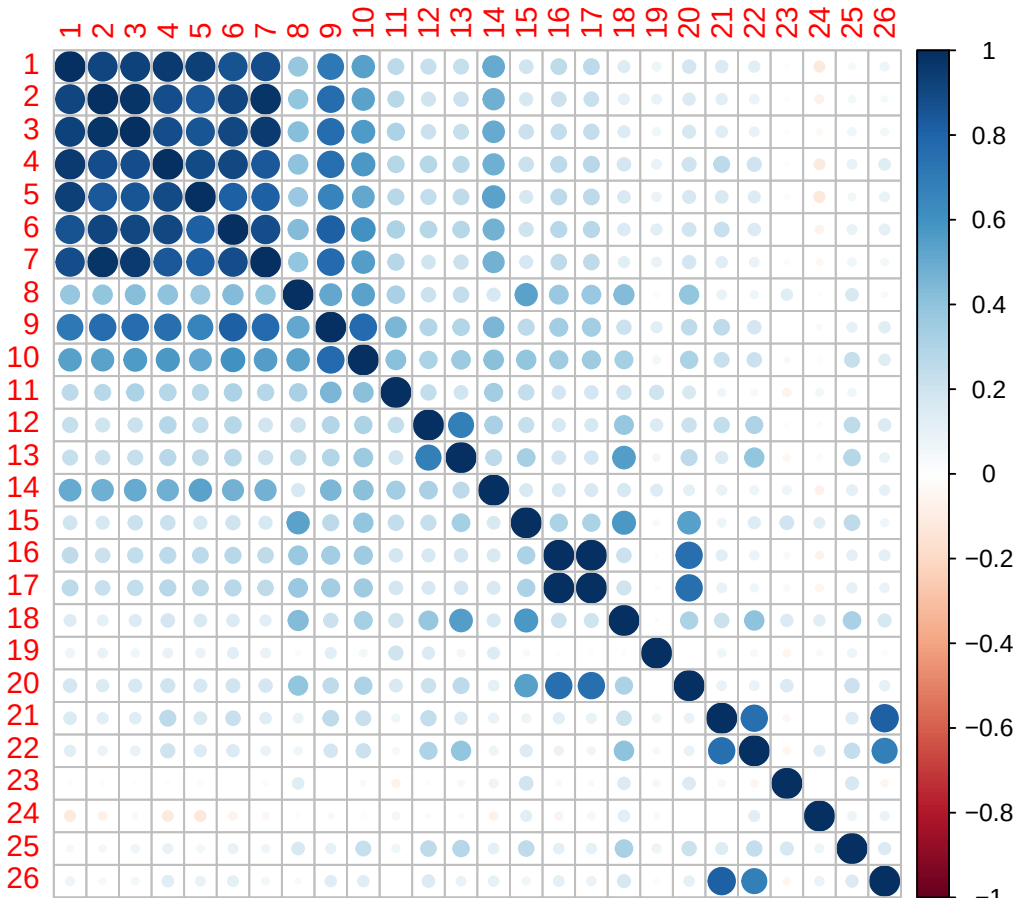
```
## Warning in abbreviate(rownames(r), minlength = 5): abbreviate used with
## non-ASCII chars
```

```
# Draw table with t-tests
corr_log <- as.data.frame(test_holm$r)
corr_test_log <- as.data.frame(test_holm$p)
order_asc_log <- order(corr_log, decreasing = TRUE)
names_log <- row.names(corr_log)[order_asc_log]
corr_log_matrix <- cor(count_select[, order_asc_log])
corr_test_log <- round(corr_test_log[order_asc_log, ], digits = 4)
corr_log <- round(corr_log[order_asc_log, ], digits = 4)
colnames(corr_log_matrix) <- seq(NCOL(corr_log_matrix))
summary_corr <- as.data.frame(cbind(names_log, colnames(corr_log_matrix),
    corr_log, corr_test_log))
colnames(summary_corr) <- c("Description", "Index", "Coefficient", "p-value")
knitr::kable(summary_corr, align = c("l"))
```

Description	Index	Coefficient	p-value
Test - has submitted the attempt with	1	0.3517	0
Test - has started the attempt with	2	0.3487	0
Test - has viewed the summary for the attempt with	3	0.336	0
Test - has had their attempt with	4	0.3348	0
Kernsystem - updated the grade with	5	0.3229	0
Test - viewed the quiz activity with course module	6	0.3189	0
Test - has viewed the attempt with	7	0.2951	0
Textseite - viewed the page activity with course module	8	0.2817	1e-04
Kernsystem - viewed the section number of the course with	9	0.2763	2e-04
Kernsystem - viewed the course with	10	0.1853	0.0602
Datei - viewed the resource activity with course module	11	0.1756	0.092
Forum - viewed the forum activity with course module	12	0.1664	0.1341
Forum - has viewed the discussion with	13	0.1467	0.2989
Kernsystem - updated the completion state for the course module with	14	0.1367	0.4178
Blog - viewed the oublog activity with course module	15	0.13	0.5004

Description	Index	Coefficient	p-value
Kernsystem - added the user with	16	0.1006	1
Gruppenwahl - Der Nutzer hat in der Gruppenwahl eine Gruppe auswÃhlt	17	0.099	1
Forum - viewed the instance list for the module forum in the course with	18	0.0924	1
Link/URL - viewed the url activity with course module	19	0.0884	1
Gruppenwahl - viewed the choicegroup activity with course module	20	-0.014	1
Forum - has created the discussion with	21	-0.0249	1
Forum - has posted content in the forum post with	22	-0.0298	1
Blog - viewed post on the oublog with the course module	23	-0.0385	1
Kernsystem - viewed the list of users in the course with	24	-0.0457	1
Kernsystem - viewed the profile for the user with	25	-0.0906	1
Forum - subscribed the user with	26	-0.1135	0.8313

```
# Plot correlations
colnames(corr_log_matrix) <- seq(NCOL(corr_log_matrix))
rownames(corr_log_matrix) <- seq(NCOL(corr_log_matrix))
corrplot(corr_log_matrix)
```



We can also produce plots. The corresponding outputs are automatically inserted in our output document. We use the built-in dataset `cars`.

```
par(mar = c(4, 4, 0.1, 0.1))
with(mtcars, {
    plot(mpg ~ hp, pch = 20, cex = 2, col = "darkgray")
    lines(lowess(hp, mpg))
```

```
})
```

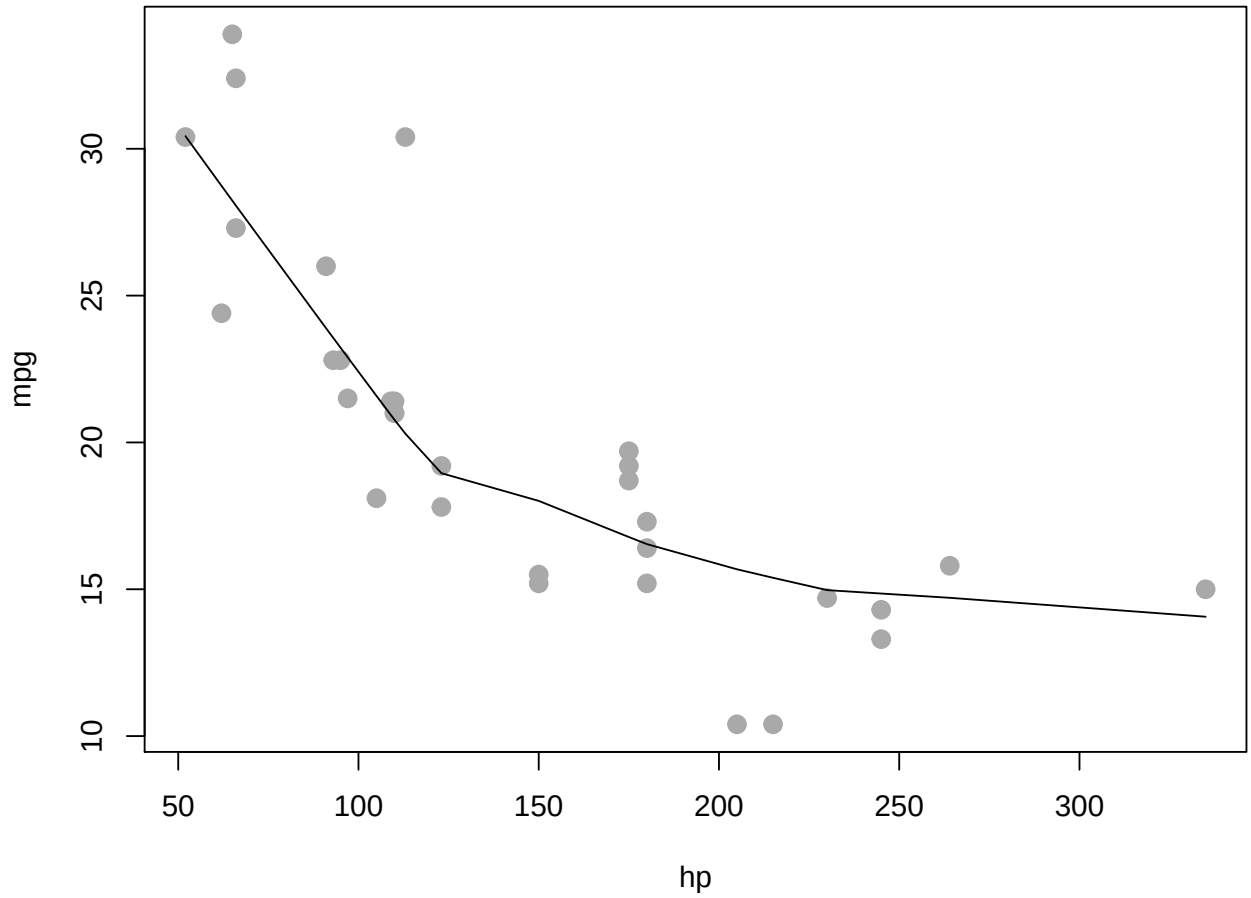


Figure 1: Your nice caption here.

```
x <- 2
```

In Figure 1 we can observe that ...

Inline Code

So called **inline R code** is also supported, e.g. the value of `x` is 2 and today is Wed May 17 14:08:59 2017. What about a rounded random number, like 1.19? Wait, why doesn't it change?

Math

LaTeX math can be used as usual: $f(\alpha, \beta) \propto x^{\alpha-1}(1-x)^{\beta-1}$. We can also write

$$2^2 + y^2 = 1.$$

Misc

Indenting

You can indent code chunks so they can nest within other environments such as lists.

1. the area of a circle with radius x

   ```
   2^2 * pi
   ```

   ```
   ## [1] 12.57
   ```

2. OK, that is great!

Tables

We can also produce pretty tables of data frames

```
knitr::kable(swiss[1:10, ], caption = "The first 10 rows of the `swiss` dataset.")
```

Table 3: The first 10 rows of the `swiss` dataset.

	Fertility	Agriculture	Examination	Education	Catholic	Infant.Mortality
Courtelary	80.2	17.0	15	12	9.96	22.2
Delemont	83.1	45.1	6	9	84.84	22.2
Franches-Mnt	92.5	39.7	5	5	93.40	20.2
Moutier	85.8	36.5	12	7	33.77	20.3
Neuveville	76.9	43.5	17	15	5.16	20.6
Porrentruy	76.1	35.3	9	7	90.57	26.6
Broye	83.8	70.2	16	7	92.85	23.6
Glane	92.4	67.8	14	8	97.16	24.9
Gruyere	82.4	53.3	12	7	97.67	21.0
Sarine	82.9	45.2	16	13	91.38	24.4

Alternatively, we can build a table ourselves manually (here with centered columns).

ETH Zurich	ETH Lausanne
-1.6896, 1.2395, -0.109	Cell 2
Cell 3	Cell 4

Results

Conclusion

Recommendations

Acknowledgments

Conclusions

Markdown is super easy to write. Go to **knitr** homepage for more details.