Cognitive Aspects of Software Development: a SMS

```
# Keep things nice and tidy, all libraries go here
library(readxl)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.2.1     v purrr   0.3.3
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   1.0.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
```

```
## -- Conflicts ------------------------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(knitr)
library(kableExtra)
```

```
##
## Attaching package: 'kableExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##     group_rows
```

```
library(svglite)
```

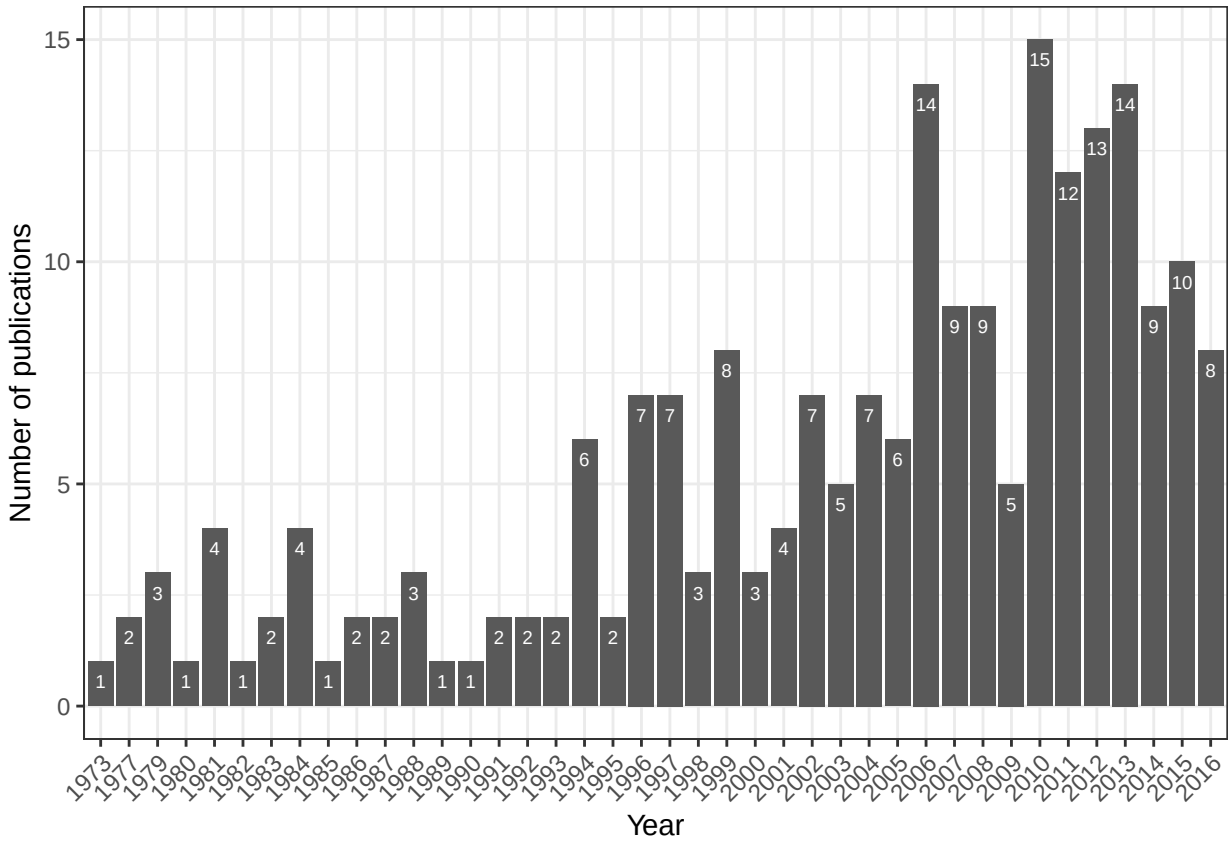
```
data <- read_excel("data.xlsx", skip = 1)
```

```
## New names:
## * `` -> ...35
## * `` -> ...71
```

```
data <- data %>% filter(is.na(Exclude))
```

Visualizing number of publications over time

```
ggplot(data, aes(x=as.factor(Year))) +
  geom_bar() +
  ylab("Number of publications") +
  xlab("Year") +
  geom_text(stat='count', aes(label=..count..), vjust=2, color="white", size = 2.5) +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```
ggsave("yearly_distribution.eps")
```

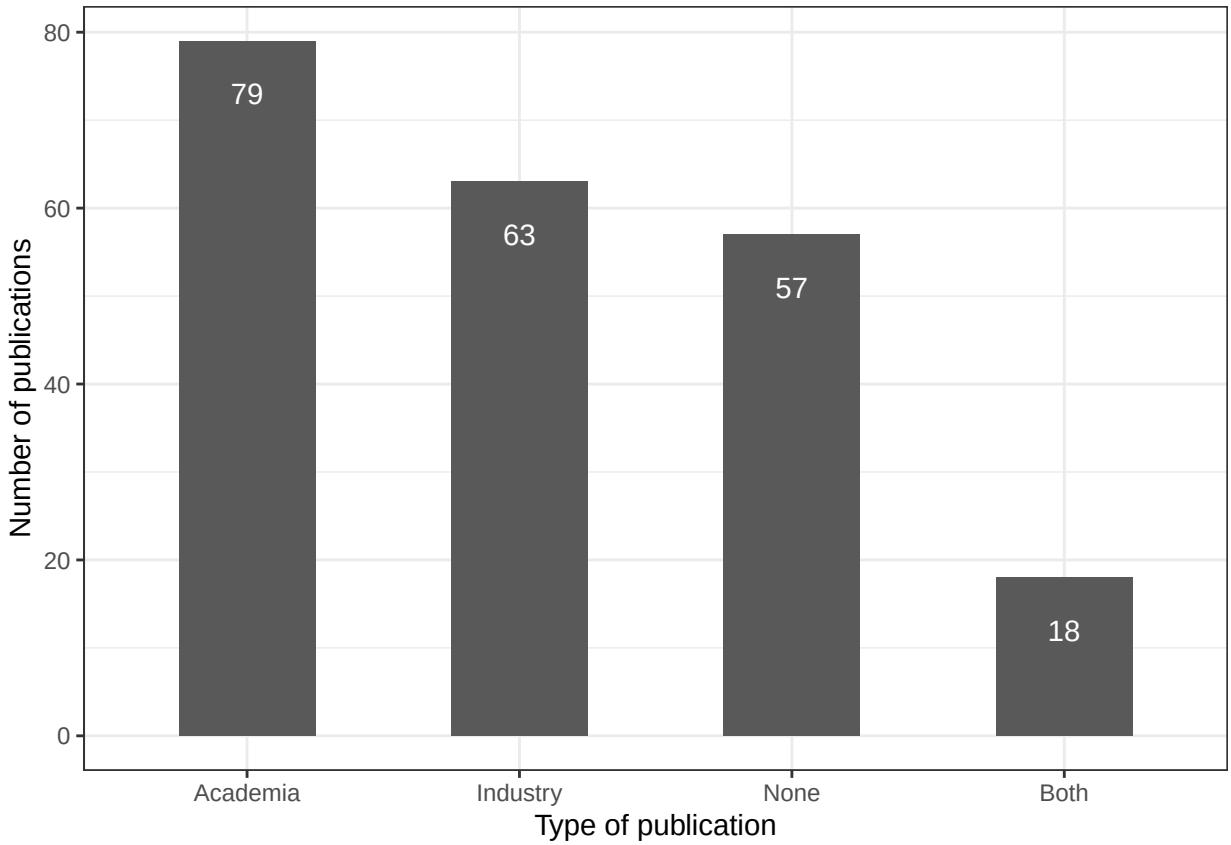
```
## Saving 6.5 x 4.5 in image
```

```
# Cleaning not needed values
data<-data %>%
    mutate(Academia = replace(Academia, Academia == "?", NA)) %>%
    mutate(Industry = replace(Industry, Industry == "?", NA))

data<-data %>%
  mutate(Type = case_when(is.na(Academia) & is.na(Industry) ~ "None",
                          Academia == "1.0" & is.na(Industry) ~ "Academia",
                          Industry == "1.0" & is.na(Academia) ~ "Industry",
                          TRUE ~ "Both"))
```

Number of publications according to their type

```
data %>%
  mutate(Type = fct_infreq(Type, ordered = T)) %>%
ggplot(aes(x=Type)) +
  geom_bar(width = .5) +
  xlab("Type of publication") +
  ylab("Number of publications") +
  geom_text(stat='count', aes(label=..count..), vjust=3, color="white", size = 4) +
  theme_bw()
```



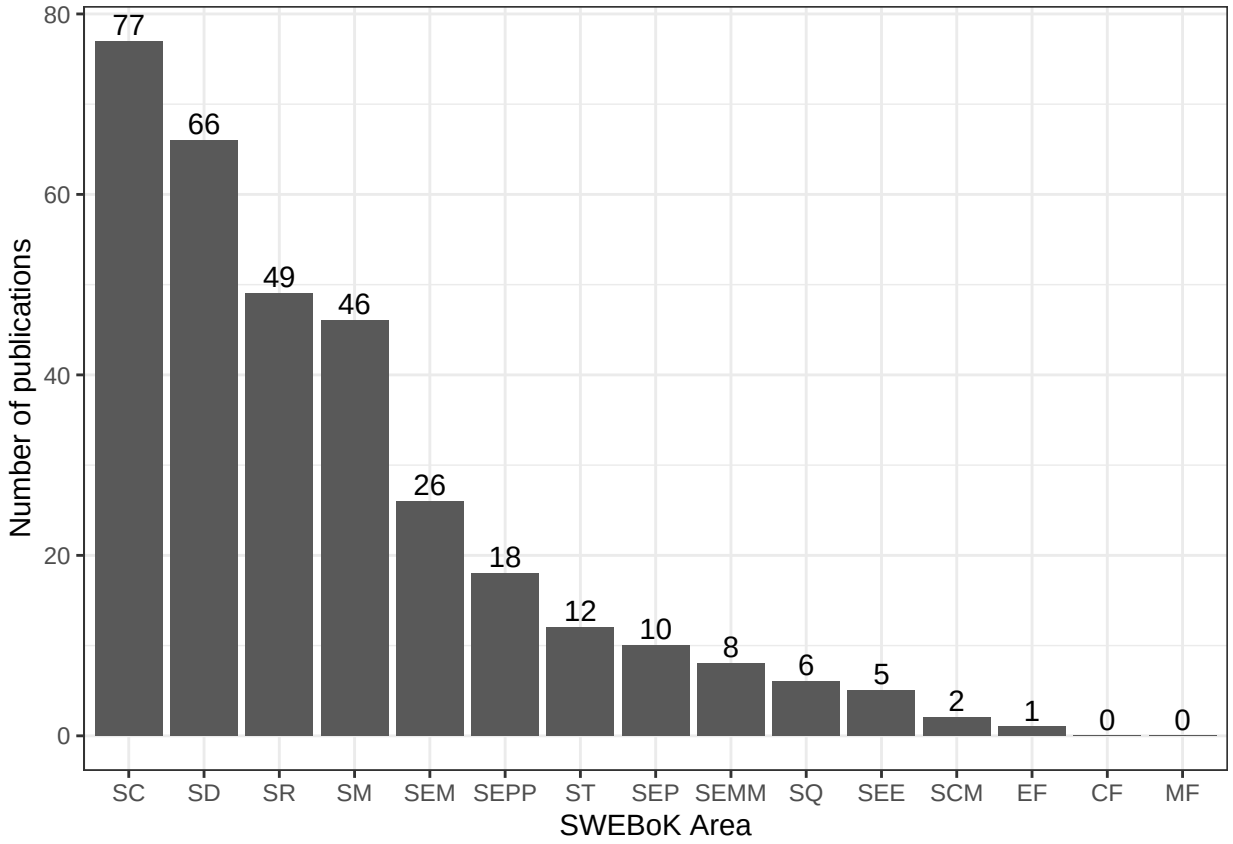

```
ggsave("academia_industry_distribution.eps")
```

```
## Saving 6.5 x 4.5 in image
```

Number of publications categorized according to SWEBoK Areas.

A publication can be in more than one category at the same time.

```
data %>%
  select(7:21) %>% # selecting columns corresponding to the SWEBoK Areas
  mutate_all(replace_na,0) %>%
  summarise_all(sum) %>%
  gather(key = "SWEBOKArea", value = "publications", 1:15) %>%
  arrange(-publications) %>%
  mutate(SWEBOKArea = factor(SWEBOKArea, SWEBOKArea)) %>%
  ggplot(aes(x=SWEBOKArea, y=publications)) +
  geom_bar(stat="identity") +
  geom_text(aes(label=publications), vjust=-0.3, color="black", size = 4) +
  xlab("SWEBoK Area") +
  ylab("Number of publications") +
  theme_bw()
```




```
ggsave("swebok_distribution.eps")
```

```
## Saving 6.5 x 4.5 in image
```

Co-occurrences of SWEBoK Areas

```
data %>%
  select(7:21) %>% # selecting columns corresponding to the SWEBoK Areas
  mutate_all(replace_na,0) %>%
  as.matrix() %>%
  crossprod() %>%
  kable() %>%
  kable_styling("striped", full_width = T) %>%
  column_spec(1, bold = T)
```

	SR	SD	SC	ST	SM	SCM	SEM	SEP	SEMM	SQ	SEPP	SEE	CF	MF	EF
SR	49	18	5	2	4	0	7	2	4	0	7	0	0	0	1
SD	18	66	17	3	4	0	6	2	6	1	6	0	0	0	1
SC	5	17	77	5	22	1	3	2	2	0	3	0	0	0	0
ST	2	3	5	12	4	0	1	0	0	0	0	0	0	0	0
SM	4	4	22	4	46	1	2	1	0	0	1	0	0	0	0
SCM	0	0	1	0	1	2	0	1	0	0	0	0	0	0	0
SEM	7	6	3	1	2	0	26	3	1	0	7	3	0	0	1
SEP	2	2	2	0	1	1	3	10	0	0	2	1	0	0	0
SEMM	4	6	2	0	0	0	1	0	8	0	1	0	0	0	0
SQ	0	1	0	0	0	0	0	0	0	6	0	0	0	0	0
SEPP	7	6	3	0	1	0	7	2	1	0	18	3	0	0	1
SEE	0	0	0	0	0	0	3	1	0	0	3	5	0	0	0
CF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EF	1	1	0	0	0	0	1	0	0	0	1	0	0	0	1

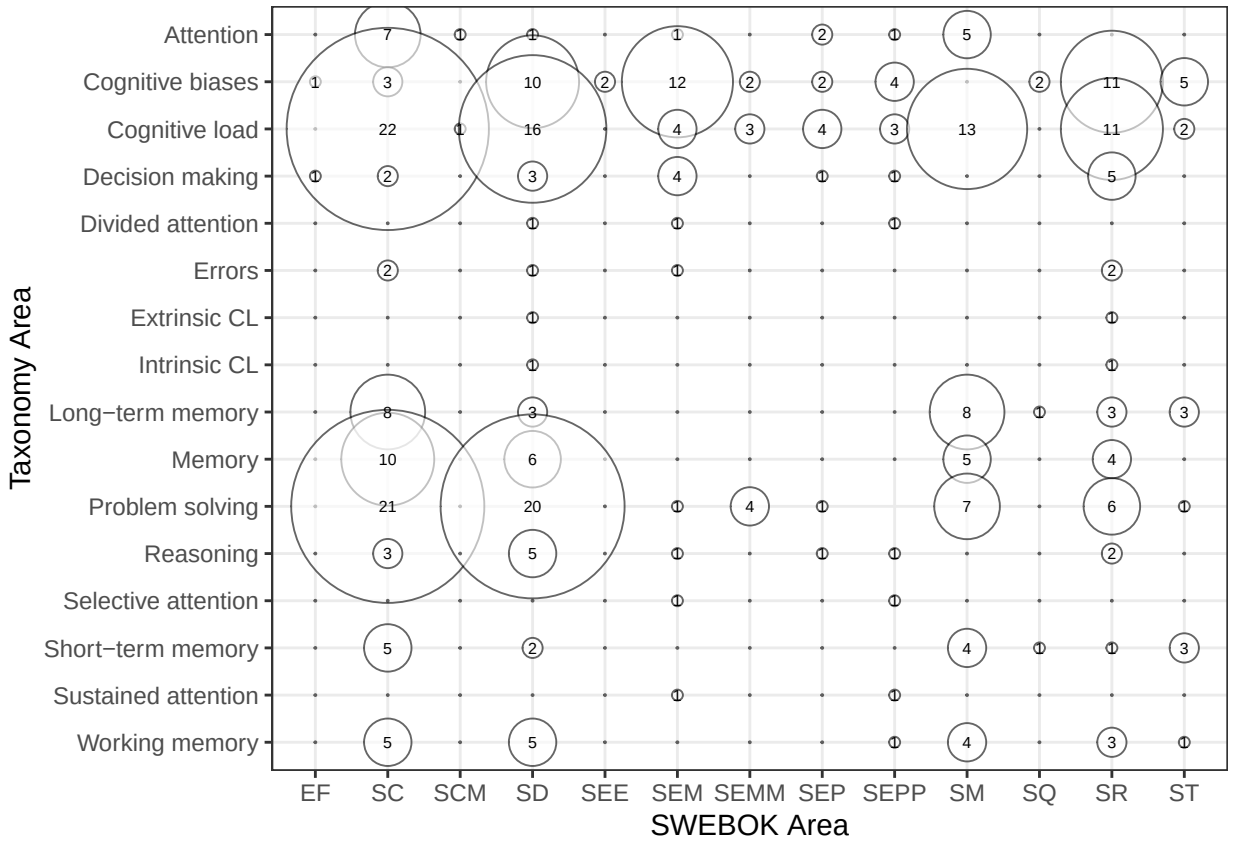
```
x <- data %>% select(7:21, matches('Attention|Memory|Cognitive load|CL$|Problem solving|Reasoning|Decisi
  mutate_all(replace_na, 0) %>%
  mutate(`Problem solving`, `Problem solving` = as.numeric(`Problem solving`)) %>%
  gather(key="SWEBOK", value = pubs, 1:15) %>% # use SWEBOK area as factor
  filter(pubs == 1) %>% # select areas for which there are publications
  group_by(SWEBOK) %>%
  summarise_all(sum) %>% # number of publication for each area
  select(-pubs) %>%  # remove pubs to reuse it later
  gather(key = "Taxonomy", value = "count", 2:17) %>%  # count publications in each cognitive taxonomy
  mutate(label = str_replace(as.character(count), "^0", "")) # add label for later
```

```
## Warning: NAs introduced by coercion
```

```
# Bubble plot
x <- arrange(x, Taxonomy)
xf<-x$Taxonomy
xfu<-unique(xf)
x$Taxonomy<-factor(xf,levels=xfu)
p<-ggplot(x)
p + geom_point(aes(x = fct_infreq(SWEBOK), y = fct_rev(Taxonomy)), size=x$count*1.6, shape=21, fill="whi
geom_text(aes(x = fct_infreq(SWEBOK), y = fct_rev(Taxonomy), label=label), size=2) +
theme(axis.text.x = element_text(angle = 45, hjust = 1.1, size=9,colour="black"), axis.text.y = element_
  labs(x="SWEBOK Area",y = "Taxonomy Area") + theme_bw()
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

```
## Warning: Removed 1 rows containing missing values (geom_text).
```




```
ggsave("swebok_taxonomy_bubble.eps")
```

```
## Saving 6.5 x 4.5 in image
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

```
## Warning: Removed 1 rows containing missing values (geom_text).
```

```
## Warning in grid.Call.graphics(C_points, x$x, x$y, x$pch, x$size): semi-
## transparency is not supported on this device: reported only once per page
```

```
# Preparing the dataset for analysing the research methods
data<-data %>%
  mutate(Quantitative = case_when(`Quantit. measures` == 1 | `Task performance` == 1 | `Physiological m
  mutate(Quantitative = replace_na(Quantitative, 0)) %>%
  mutate(Qualitative = case_when(Fieldwork == 1 | Interview == 1 | `Qualit. measures` == 1 | `Task-base
  mutate(Qualitative = replace_na(Qualitative, 0)) %>%
  mutate(Both = if_else(Qualitative == 1 & Quantitative == 1, 1, 0))
```