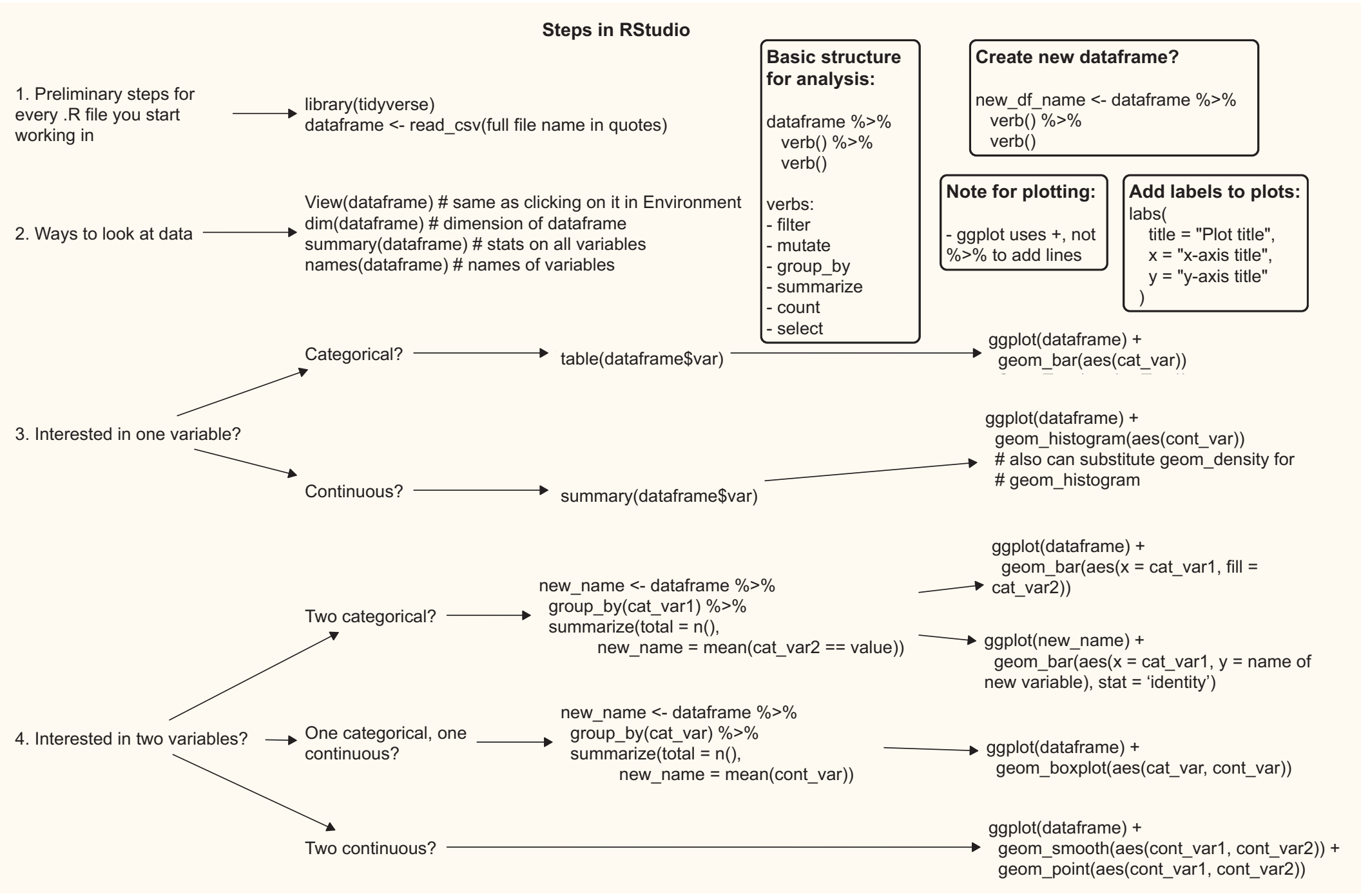
Steps in RStudio
1. Preliminary steps for every .R file you start working in
library(tidyverse)
dataframe <- read_csv(full file name in quotes)
Basic structure for analysis:
dataframe %>%
verb() %>%
verb()
verbs:
- filter
- mutate
- group_by
- summarize
- count
- select
Create new dataframe?
new_df_name <- dataframe %>%
verb() %>%
verb()
Note for plotting:
- ggplot uses +, not %>% to add lines
Add labels to plots:
labs(
title = "Plot title",
x = "x-axis title",
y = "y-axis title"
)
2. Ways to look at data
View(dataframe) # same as clicking on it in Environment
dim(dataframe) # dimension of dataframe
summary(dataframe) # stats on all variables
names(dataframe) # names of variables
3. Interested in one variable?
Categorical?
table(dataframe$var)
ggplot(dataframe) +
geom_bar(aes(cat_var))
Continuous?
summary(dataframe$var)
ggplot(dataframe) +
geom_histogram(aes(cont_var))
# also can substitute geom_density for
# geom_histogram
4. Interested in two variables?
Two categorical?
new_name <- dataframe %>%
group_by(cat_var1) %>%
summarize(total = n(),
new_name = mean(cat_var2 == value))
ggplot(dataframe) +
geom_bar(aes(x = cat_var1, fill = cat_var2))
ggplot(new_name) +
geom_bar(aes(x = cat_var1, y = name of new variable), stat = 'identity')
One categorical, one continuous?
new_name <- dataframe %>%
group_by(cat_var) %>%
summarize(total = n(),
new_name = mean(cont_var))
ggplot(dataframe) +
geom_boxplot(aes(cat_var, cont_var))
Two continuous?
ggplot(dataframe) +
geom_smooth(aes(cont_var1, cont_var2)) +
geom_point(aes(cont_var1, cont_var2))