

### Read / Write .csv

```
read.csv(file.csv)
write.csv(df, file.csv)
```

### Meta Data

```
summary(df)          str(df)
ls()                 dir()
length(df)           dim(df)
class(df)            nrow(df) / ncol(df)
colnames(df)
```

### Arrange

```
rename(df, varold = varnew)
select(df, var1, var2)
arrange(df, var) / arrange(df, desc(var))
```

### Filter

```
filter(df, var1 > 10) / df[df$var1 > 10, ]
slice(df, 10:15) / df[10:15, ] # select rows
distinct(df, var)
sample_frac(df, 0.5, replace = FALSE) / sample_n(df,
10, replace = FALSE)
select(df, col1, col2, ...) / df[, c('col1', 'col2')]
select(df, contains() starts_with() ends_with())
na.omit(df)
```

### Useful Functions

```
min() / max()          first() / last()
n() / n_distinct()     mean() / median()
lead() / lag()         var() / sd()
between(a, b)          cumsum()
is.na()                %in%
if_else() / case_when() coalesce()
na_if()                sum(!is.na())
quantile(x, probs=)    seq(1, 10)
rep(c(1,2,3), 2)       is.na() / is.null() /
                       is.numeric()
as.character() /       replace(x, list, values)
as.numeric() ...
round(x, n)            mod()
```

### Useful Functions (cont)

```
abs()                  corr(x) / corr(x, y)
weighted.mean(x, w)    paste(vect, sep) /paste0()
substr(x, 1, 3)        grep(pat, repl, x)
tolower() / toupper() nchar(x)
```

### Write Functions

```
function_name <- function(x, y) {
  if (statement) {
    do
  } elif (statement) {
    do
  }
  for (year in 201:2015) {
    do
  }
  return(result)
}
```

### Applying Functions

```
mutate_each(df, funs(sum))
summarise_each(df, funs(sum))
apply(x, index, fun)
lapply(x, fun) / sapply(x, fun)
```

### Summarise

```
group_by(df, var) %>%
  summarise(avg = mean(val))
  count(var, wt = weight)
table(df$var)
aggregate(x, by, fun)
```

### Join

```
left_join(df1, df2, by = "var") / right_join(df1, df2,
by = "var")
inner_join(df1, df2, by = "var") / full_join(df1, df2,
by = "var")
anti_join(df1, df2, by = "var") / semi_join(df1, df2,
by = "var")
```



### Method Chaining

```
df = df %>%  
  select(var1, var2) %>%  
  mutate(newvar = var1 + var2)
```

### New Variable / Column

```
df$newvar = df$var1 + df$var2  
mutate(df, newvar = var1 + var2)  
transmute(df, newvar = var1 + var2) # drop orig cols
```

### Reshape Data

```
gather(df, 'var', 'val') # columns into rows  
spread(df, var, val) # rows into columns  
unite(df, col1, col2, sep) # sev cols into 1  
bind_rows(df1, df2) / bind_cols(df1, df2) / rbind(...)  
/ cbind(..)  
melt() / cast() / recast() / reshape()
```



By **mitch**  
[cheatography.com/mitch/](http://cheatography.com/mitch/)

Not published yet.  
Last updated 29th December, 2017.  
Page 2 of 2.

Sponsored by **CrosswordCheats.com**  
Learn to solve cryptic crosswords!  
<http://crosswordcheats.com>