

# Reference sheet

## dplyr

- `select()` select columns
- `filter()` filter rows
- `group_by()` group data by category
- `summarize()` summarize data using functions of choice
- `arrange()` arrange rows
- `mutate()` create new columns
- `rename()` rename columns

## dplyr functions

- `n()` (inside `summarize()`/'mutate/filter only)
- `n_distinct()`

## vectors

- `unique()`
- `sort()`
- `length()`
- `1:10`
- `seq(-10, 10, 0.1)`

## sapply

- `sapply(X = v, FUN = f, extra1, extra2, extra3, ...)`

## grep

- `grep(pattern, text)[[1]]`

## classes

- `numeric`
- `character`
- `factor`
- `logical`

## data.frame

- `my.data[row, column]`
- `data.frame(col1 = col1.vec, col2 = col2.vec)`

## Regression

- `lm(formula, data = my.data)`
- `glm(formula, data = my.data, family = "binomial")`
- `predict(fit, newdata = my.data)`
- `predict(fit, newdata = my.data, type = "response")`
- `fit$coefficients`

## Conditionals

```
if(<COND1>){
  ...
} else if(<COND2>){
  ...
} else {
  ...
}
```

## ggplot

```
ggplot(data = gapminder,
        mapping = aes(x = gdpPercap, y = lifeExp)) +
  geom_point() +
  geom_smooth(method = "loess") # or method = "lm"
```

## geom\_bar

- `aes(y = ..count..)`
- `aes(y = ..prop..)`
- `aes(y = ..prop.., group = 1)`
- `aes(y = ..prop.., group = VAR)`
- position: `stack` (default), `identity`, `dodge`
- stat: `count` (default), `identity`, `bin` (same as `geom_histogram`)