* Pick observations by their values (filter()).
* Reorder the rows (arrange()).
* Pick variables by their names (select()).
* Create new variables with functions of existing variables (mutate()).
* Collapse many values down to a single summary (summarise()).

**filter**(flights, !(arr\_delay > 120 | dep\_delay > 120))

**filter**(flights, arr\_delay <= 120, dep\_delay <= 120)

**select**(flights, year, month, day)

*#> # A tibble: 336,776 x 3*

*#> year month day*

*#> <int> <int> <int>*

*#> 1 2013 1 1*

*#> 2 2013 1 1*

*#> 3 2013 1 1*

*#> 4 2013 1 1*

*#> 5 2013 1 1*

*#> 6 2013 1 1*

*#> # … with 3.368e+05 more rows*

*# Select all columns between year and day (inclusive)*

flights\_sml <- **select**(flights,

year:day,

**ends\_with**("delay"),

distance,

air\_time

)

**mutate**(flights\_sml,

gain = dep\_delay - arr\_delay,

speed = distance / air\_time \* 60

)

*#> # A tibble: 336,776 x 9*

*#> year month day dep\_delay arr\_delay distance air\_time gain speed*

*#> <int> <int> <int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>*

*#> 1 2013 1 1 2 11 1400 227 -9 370.*

*#> 2 2013 1 1 4 20 1416 227 -16 374.*

*#> 3 2013 1 1 2 33 1089 160 -31 408.*

*#> 4 2013 1 1 -1 -18 1576 183 17 517.*

*#> 5 2013 1 1 -6 -25 762 116 19 394.*

*#> 6 2013 1 1 -4 12 719 150 -16 288.*

*#> # … with 3.368e+05 more rows*

**summarise**(flights, delay = **mean**(dep\_delay, na.rm = TRUE))

*#> # A tibble: 1 x 1*

*#> delay*

*#> <dbl>*

*#> 1 12.6*

by\_dest <- **group\_by**(flights, dest)

delay <- **summarise**(by\_dest,

count = **n**(),

dist = **mean**(distance, na.rm = TRUE),

delay = **mean**(arr\_delay, na.rm = TRUE)

)

delay <- **filter**(delay, count > 20, dest != "HNL")

*# It looks like delays increase with distance up to ~750 miles*

*# and then decrease. Maybe as flights get longer there's more*

*# ability to make up delays in the air?*

**ggplot**(data = delay, mapping = **aes**(x = dist, y = delay)) +

**geom\_point**(**aes**(size = count), alpha = 1/3) +

**geom\_smooth**(se = FALSE)

*#> `geom\_smooth()` using method = 'loess' and formula 'y ~ x'*