

Data wrangling verbs

*in the **dplyr** package*

select()
filter()
summarize()
group_by()
mutate()
arrange()



& the pipe operator

$\%>\%$



`select()`

Pull out just the **columns** you want in a data set, based on the column names

Subset Variables (Columns)



select()

Subset Variables (Columns)



Example

```
flights_sub <- flights %>%  
  select(arr_delay, dep_delay)
```

the data
frame

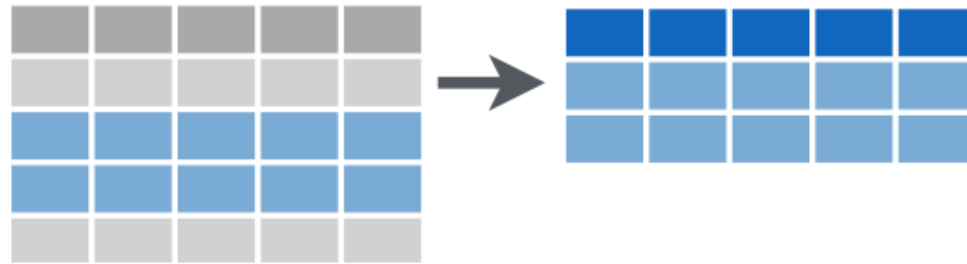
the verb
(function)

what columns (variable)
to keep

filter()

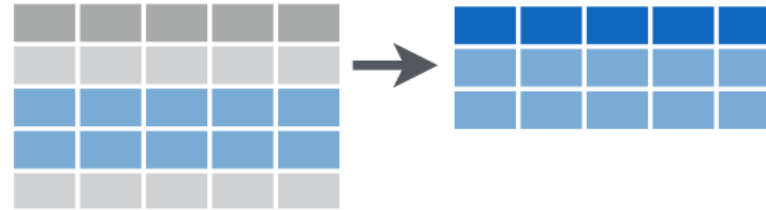
Pull out just the **rows** you want in a data set, based on some criteria

Subset Observations (Rows)



filter()

Subset Observations (Rows)



Example

```
on_time_flights <- flights %>%  
  subset(arr_delay < 30)
```

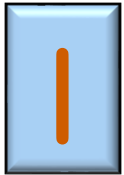
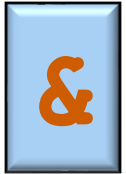
the data
frame

Only keep rows of the
data in which the column
meets this criteria

the verb
(function)

what column (variable)
to use for filtering

filter()



Another
example



Subset Observations (Rows)



Only keep rows of the
data in which the column
meets this criteria

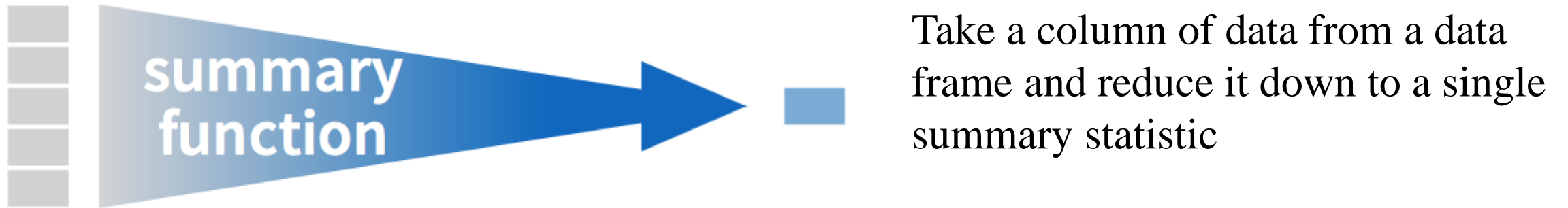
```
on_time_flights <- flights %>%  
  subset(origin == "AK")
```

the data
frame

the verb
(function)

what column (variable)
to use for filtering

summarize()



Summarise Data



summarize()



Example

the data frame

```
flights %>%  
  summarize(av_dist = mean(distance))
```

what summary
statistic to
calculate

the verb
(function)

what to call the
result

the column to
summarize

`group_by()`

then

`summarize()`

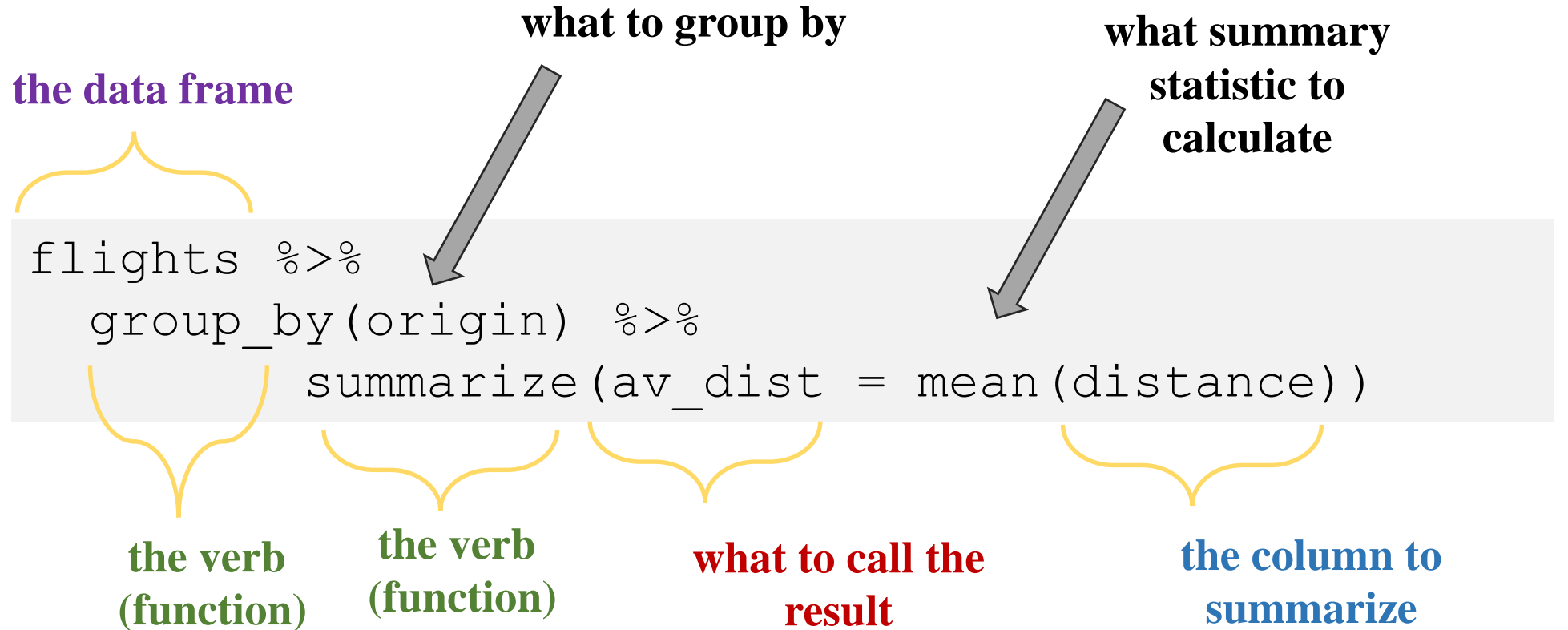


Take a column of data and reduce it
down to a summary statistic, by
some grouping variable

summarize()



Example



mutate()

Make New Variables



Mutant growth on a tomato

mutate()

Make New Variables



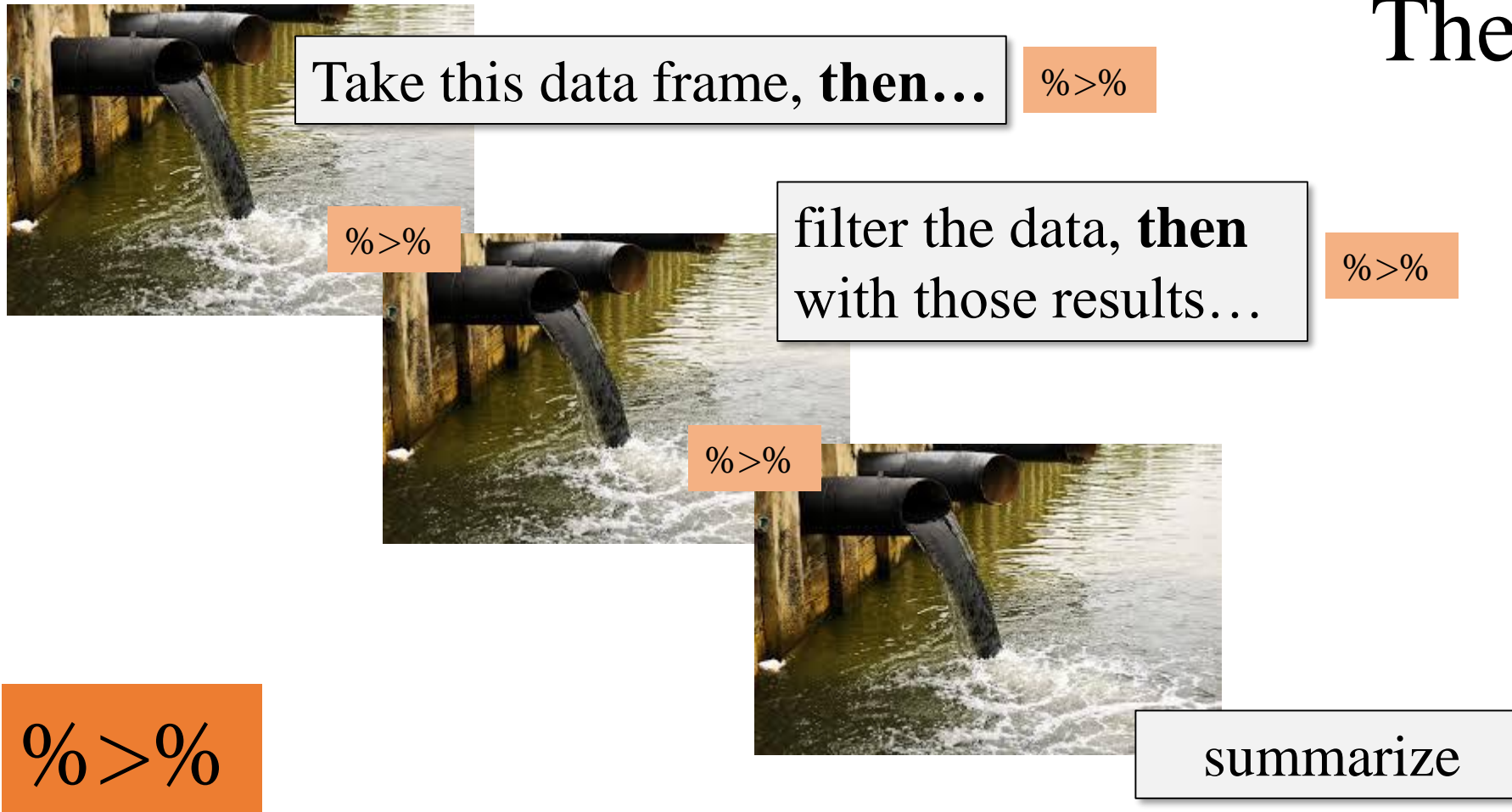
Example

the data frame **what to call the new column (new variable)**

```
flights %>%  
  mutate(dist_ft = distance * 5280)
```

the verb (function) **what to put in the new column**

The pipe operator



- “Pipe” a data frame into a “verb” command
- “Chain” the results from one “verb” command into another
- Think of it as the word “**then**”