ETC1010

Sway

Lobby | Rooms | Quizzes | Reports

Edit Quiz

Save and Exit

Quiz name

Reading week 3 (DUE: Tue Aug 15 9am)

1 **MULTIPLE CHOICE** Edit

Which of these are wrangling verbs?

Answer choices

A	filter
B	mutate
C	summarise
D	select
E	arrange

2 MULTIPLE CHOICE

Edit

The verb "filter" means to choose observations while the verb "select" means to choose _____

Answer choices

A	cases
B	subjects
C	people
D	variables

3 MULTIPLE CHOICE

Edit

The verb "mutate" adds new columns that are functions of _____ columns.

Answer choices

A	existing
B	new
C	tidy
D	half

4 **MULTIPLE CHOICE** Edit

The verb "summarise" collapses a column to a single row. It is most useful when combined with _____ to compute statistics on subsets of rows.

Answer choices

A	mutate
B	group_by
C	arrange
D	select

5 **MULTIPLE CHOICE** Edit

For this tibble:

```
> glimpse(french_fries)
Observations: 696
Variables: 9
 time      <fctr> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1…
 treatment <fctr> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1…
 subject   <fctr> 3, 3, 10, 10, 15, 15, 16, 16, 19, 19, 31, 31, 51, 51, 5…
 rep       <dbl> 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1,…
 potato    <dbl> 2.9, 14.0, 11.0, 9.9, 1.2, 8.8, 9.0, 8.2, 7.0, 13.0, 12.…
 buttery   <dbl> 0.0, 0.0, 6.4, 5.9, 0.1, 3.0, 2.6, 4.4, 3.2, 0.0, 0.0, 2…
 grassy    <dbl> 0.0, 0.0, 0.0, 2.9, 0.0, 3.6, 0.4, 0.3, 0.0, 3.1, 0.0, 0…
 rancid    <dbl> 0.0, 1.1, 0.0, 2.2, 1.1, 1.5, 0.1, 1.4, 4.9, 4.3, 2.5, 5…
 painty    <dbl> 5.5, 0.0, 0.0, 0.0, 5.1, 2.3, 0.2, 4.0, 3.2, 10.3, 2.0, …
```

using the code:

```
french_fries %>% select(-potato, -buttery, -grassy, -rancid)
```

will leave what variables in the data?

Answer choices

A	potato, buttery, grassy, rancid
B	time, treatment, subject, rep, painty
C	time, treatment, subject, potato, grassy
D	potato, buttery, grassy, rancid, painty

6 **MULTIPLE CHOICE** Edit

There are three steps to prepare some data:

1. Group flights by destination.
2. Summarise to compute distance, average delay, and number of flights.
3. Filter to remove noisy points and Honolulu airport, which is almost twice as far away as the next closest airport.

The code to do this is:

```
delays <- flights %>%
  group_by(dest) %>%
  summarise(
    count = n(),
    dist = mean(distance, na.rm = TRUE),
    delay = mean(arr_delay, na.rm = TRUE)
  ) %>%
  filter(count > 20, dest != "HNL")
```

Step 1 uses the verb _____, step 2 uses the verb _____ and step 3 uses the verb _____

Answer choices

A	filter, summarise, group_by
B	filter, group_by, summarise
C	group_by, summarise, filter
D	group_by, filter, summarise
E	summarise, group_by, filter
F	summarise, filter, group_by

7 MULTIPLE CHOICE

Edit

If you only want to keep the new variables, use _____

Answer choices

A	transmute()
B	mutate()
C	select()
D	group_by()

8 TRUE/FALSE

Edit

Use desc() to re-order by a column in descending order, for example:

arrange(flights, desc(arr_delay))

Answer

True

9 MULTIPLE CHOICE

Edit

In the code:

```
flights %>%
  group_by(year, month, day) %>%
  summarise(mean = mean(dep_delay, na.rm = TRUE))
```

what would changing the argument "na.rm = FALSE" do? (Choose all that apply)

Answer choices

A	Keeping the missing values in the calculation
B	Return an NA for the mean if missing values exist
C	Ignore the missing values for the calculation
D	Do the calculation to the na room

10 MULTIPLE CHOICE

Edit

How can you tell whether an object is a tibble? (Check all that apply)

Answer choices

A	is.data.frame()
B	is.tibble()
C	is_tibble()
D	typeof()

11 | **MULTIPLE CHOICE** Edit

In the data:

```
> N4YRAA_latlon
# A tibble: 146 x 11
     FL_DATE CARRIER FL_NUM ORIGIN  DEST DEP_TIME ARR_TIME DISTANCE
      <date>   <chr>  <int>  <chr> <chr>    <chr>    <chr>    <dbl>
 1 2017-05-26      AA   2246    CVG   DFW     0748     0917      812
 2 2017-05-02      AA   2276    DFW   IND     2020     2323      761
 3 2017-05-05      AA   2278    DFW   OKC     0848     0941      175
 4 2017-05-11      AA   2287    STL   ORD     0454     0600      258
...
```

which variables are consider numeric according to this summary? (Choose all that apply)

Answer choices

A	FL_DATE
B	CARRIER
C	FL_NUM
D	ORIGIN
E	DEST
F	DEP_TIME
G	ARR_TIME
H	DISTANCE

12 **SHORT ANSWER** Edit

Another way to create a tibble is with tribble(), short for _____ tibble. tribble() is customised for data entry in code: column headings are defined by formulas (i.e. they start with ~), and entries are separated by commas. This makes it possible to lay out small amounts of data in easy to read form.

Answer

transposed

13 **SHORT ANSWER** Edit

The variables used to connect each pair of tables are called ____.

Answer

keys

14 **TRUE/FALSE** Edit

A mutating join allows you to combine variables from two tables

Answer

True

15 | MULTIPLE CHOICE

Edit

A left join keeps all observations in

Answer choices

A	left side table
B	right side table
C	both tables
D	neither tables

+ Multiple Choice + True/False + Short Answer