

# ds4biomed

## Pre-Workshop Exercise

### Pre-Workshop Exercise

#### Exercise 1

#### Exercise 2

#### Exercise 3

#### Summative


Start Over

Please write the code for the following pipeline steps:

1. Load the `tidyverse` and `readxl` libraries.
2. Read in the Excel file located in:  
"data/medicaldata\_tumorgrowth.xlsx" into a variable `tumor`.
3. Select the all the columns except `Grp`, and filter the rows such that `Day` is `0` or `20`. Save this data subset into a variable `tumor_subset`.
4. We want to compare baseline tumor sizes (Day 0) with tumor sizes at Day 20 between each of the groups.
  - Using `tumor_subset`, calculate the average `tumor Size` for each `Grp` and `Day`.
5. Save `tumor_subset` into a CSV file located in "data/tumorsubset.csv".

R Code

 Start Over

 Run Code

```
1  
2  
3
```

Next Topic

## ds4biomed

## Exercise 1

Pre-Workshop Exercise

Exercise 1

Exercise 2

Exercise 3

Summative

Start Over

Take a look at the `ebola` dataset.

ebola						
date	country	289	288	287	286	
<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	
1/5/2015	Guinea	2776	NA	NA	NA	
1/5/2015	SierraLeone	10030	NA	NA	NA	
1/4/2015	Guinea	NA	2775	NA	NA	
1/4/2015	SierraLeone	NA	9780	NA	NA	
1/3/2015	Guinea	NA	NA	2769	NA	
1/3/2015	Liberia	NA	NA	8166	NA	
1/3/2015	SierraLeone	NA	NA	9722	NA	
1/2/2015	Liberia	NA	NA	NA	8157	
12/31/2014	Guinea	NA	NA	NA	NA	
12/31/2014	Liberia	NA	NA	NA	NA	
1...	Previous	1	2	3	4	5 6 ... 37 Next

- Tidy the dataset such that you get the dataset below.
- You can use the `last_col()` to select the last column of the dataset.
- Remember to drop missing values as the last step.

date	country	day	cases
<chr>	<chr>	<chr>	<dbl>
1/5/2015	Guinea	289	2776
1/5/2015	SierraLeone	289	10030
1/4/2015	Guinea	288	2775
1/4/2015	SierraLeone	288	9780
1/3/2015	Guinea	287	2769
1/3/2015	Liberia	287	8166

# ds4biomed

Pre-Workshop Exercise

Exercise 1

Exercise 2

Exercise 3


Summative

Start Over

date	country	day	cases
<chr>	<chr>	<chr>	<dbl>
1/3/2015	SierraLeone	287	9722
1/2/2015	Liberia	286	8157
12/31/2014	Guinea	284	2730
12/31/2014	Liberia	284	8115
1... Previous	1 2 3 4 5 6 ... 37	Next	

R Code

 Start Over

 Run Code

```
1 |  
2  
3
```

Previous Topic

Next Topic

## ds4biomed

## Exercise 2

## Pre-Workshop Exercise

Exercise 1

Exercise 2

Exercise 3

Summative

Start Over

This is a different version of the `ebola` dataset.

date <chr>	da <dbl>	Cases_Guinea <dbl>	Cases_Liberia <dbl>
1/5/2015	289	2776	NA
1/4/2015	288	2775	NA
1/3/2015	287	2769	8166
1/2/2015	286	NA	8157
12/31/2014	284	2730	8115
12/28/2014	281	2706	8018
12/27/2014	280	2695	NA
12/24/2014	277	2630	7977
12/21/2014	273	2597	NA
12/20/2014	272	2571	7862
1... Previous	1	2	3

- Tidy the dataset such that you get the dataset below
- Remember to drop missing values as the last step

date <chr>	da <dbl>	case_death <chr>	country <chr>	val... <dbl>
1/5/2015	289	Cases	Guinea	2776
1/5/2015	289	Cases	SierraLeone	10030
1/5/2015	289	Deaths	Guinea	1786
1/5/2015	289	Deaths	SierraLeone	2977
1/4/2015	288	Cases	Guinea	2775
1/4/2015	288	Cases	SierraLeone	9780
1/4/2015	288	Deaths	Guinea	1781
1/4/2015	288	Deaths	SierraLeone	2943
1/3/2015	287	Cases	Guinea	2769
1/3/2015	287	Cases	Liberia	8166

# ds4biomed

Pre-Workshop Exercise

Exercise 1

Exercise 2


Exercise 3


Summative

Start Over

1... Previous **1** 2 3 4 5 6 ... 74 Next

R Code

 Start Over

 Run Code

```
1 |  
2  
3
```

Previous Topic

Next Topic

## ds4biomed

## Exercise 3

Pre-Workshop Exercise

Exercise 1

Exercise 2

Exercise 3

Summative

Start Over

Using the same `ebola` dataset example from the previous exercise:

<b>date</b> <chr>	<b>day</b> <dbl>	<b>Cases_Guinea</b> <dbl>	<b>Cases_Liberia</b> <dbl>
1/5/2015	289	2776	NA
1/4/2015	288	2775	NA
1/3/2015	287	2769	8166
1/2/2015	286	NA	8157
12/31/2014	284	2730	8115
12/28/2014	281	2706	8018
12/27/2014	280	2695	NA
12/24/2014	277	2630	7977
12/21/2014	273	2597	NA
12/20/2014	272	2571	7862
1... Previous	1	2	3

- Tidy the dataset such that it looks like the below example
- Remember to drop missing values as the last step

<b>date</b> <chr>	<b>day</b> <dbl>	<b>country</b> <chr>	<b>Cases</b> <dbl>	<b>Deaths</b> <dbl>
1/5/2015	289	Guinea	2776	1786
1/5/2015	289	SierraLeone	10030	2977
1/4/2015	288	Guinea	2775	1781
1/4/2015	288	SierraLeone	9780	2943
1/3/2015	287	Guinea	2769	1767
1/3/2015	287	Liberia	8166	3496
1/3/2015	287	SierraLeone	9722	2915
1/2/2015	286	Liberia	8157	3496
12/31/2014	284	Guinea	2730	1739

# ds4biomed

## Pre-Workshop Exercise

### Exercise 1

### Exercise 2

### Exercise 3


### Summative

Start Over

date	day	country	Ca...	Deaths
<chr>	<dbl>	<chr>	<dbl>	<dbl>
12/31/2014	284	Liberia	8115	3471
1... Previous	1	2	3	4
	5	6	...	37
				Next

R Code

 Start Over

 Run Code

```
1 |  
2  
3
```

Previous Topic

Next Topic

## ds4biomed

## Summative

## Pre-Workshop Exercise

Exercise 1

Exercise 2

Exercise 3

Summative

Start Over

This is the `cmv` dataset you will load:

ID	ag	prior.radiation	aK...	cm	donor_negati				
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<chr>				
1	61	0	1	1	recipient_posit				
2	62	1	5	0	recipient_nega				
3	63	0	3	0	NA				
4	33	1	2	0	recipient_posit				
5	54	0	6	0	NA				
6	55	0	2	1	NA				
7	67	0	1	0	NA				
8	51	0	2	0	NA				
9	44	1	2	1	NA				
10	59	0	4	0	recipient_nega				
1-1...	Previous	1	2	3	4	5	6	7	Next

1. Use the `readxl` library to load the `data/cmv.xlsx` into a variable, `cmv`.
2. Filter the `cmv` dataset such that only age > 65 are remaining. Save this to a variable, `cmv_subset`.
3. Save the `cmv_subset` variable to a `csv` file in `"data/cmv_subset.csv"`.

R Code

Start Over

Run Code

```
1 library(readxl)
2 read_xlsx()
3
```

4. Tidy the `cmv` dataset such that it looks like the `clean` dataset below. Save the tidy dataset into a variable, `cmv_tidy`.

ID	age	prior.radiation	aK...	cm	donor_status
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<chr>
1	61	0	1	1	donor_negativ
2	62	1	5	0	donor_negativ



# ds4biomed

Pre-Workshop Exercise

Exercise 1

Exercise 2

Exercise 3


Summative

Start Over

ID	age	prior.radiation	aK...	cmv	donor_status
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<chr>
3	63	0	3	0	donor_positive
4	33	1	2	0	donor_negative
5	54	0	6	0	donor_positive
6	55	0	2	1	donor_positive
7	67	0	1	0	donor_positive
8	51	0	2	0	donor_positive
9	44	1	2	1	donor_positive
10	59	0	4	0	donor_negative
1-1... Previous 1 2 3 4 5 6 7 Next					

R Code

 Start Over


 Run Code

```
1 |
2
3
```

5. In the `cmv_tidy` dataset, calculate the average  
age for each value of `cmv`.

R Code

 Start Over

 Run Code

```
1 |
2
3
```

Previous Topic

# ds4biomed

Pre-Workshop Exercise

---

Exercise 1

---

Exercise 2

---

Exercise 3

---

Summative

---

[Start Over](#)