

Defunciones en Colombia - Reporte

Daniel Delgado

Table of contents

1	Antes de la limpieza	1
2	Después de la limpieza	4

1 Antes de la limpieza

Inicialmente, dado que los datos se encontraban en un Excel, y las hojas de cálculo no estaban estructuradas de manera adecuada para un análisis inmediato (usando alguna herramienta de análisis de datos), al importar los datos se incluía información de celdas que no eran precisamente parte de estos (como por ejemplo, la descripción, notas, fuentes de los datos, etc.) como se puede ver a continuación:

Estadísticas														
Vitales	...2	...3	...4	...5	...6	...7	...8	...9	...10	...11	...12	...13	...14	...15
NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Semanas según estándar ISO 8601	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Años 2015 - 2021	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Semana NA	2015 Inicia	NA Fin	2016 Inicia	NA Fin	2017 Inicia	NA Fin	2018 Inicia	NA Fin	2019 Inicia	NA Fin	2020 Inicia	NA Fin	2021 Inicia	NA Fin
semana 01	4200242008423734237942737427434310143107434654347143829438354420044206													
Semana 02	4200942015423804238642744427504310843114434724347843836438424420744213													
Semana 03	4201642022423874239342751427574311543121434794348543843438494421444220													
Semana 04	4202342029423944240042758427644312243128434864349243850438564422144227													
Semana 05	4203042036424014240742765427714312943135434934349943857438634422844234													

Estadísticas Vitales	...2	...3	...4	...5	...6	...7	...8	...9	...10	...11	...12	...13	...14	...15
NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Defunciones no fetales por probable manera de muerte y sexo, según año, departamento de residencia y semana de defunción	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Años 2015 - 2021pr (Sem 1 - 35)	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	Natural	NA	NA	NA	Violencia	NA	NA	NA	En estudio	NA	NA	NA
NA	NA	NA	Total	Masculino	Femenino	Indeterminado	Total	Masculino	Femenino	Indeterminado	Total	Masculino	Femenino	Indeterminado
2015	Total	NA	1932	983	491	0	2880	2433	1446	26	1435	979	440	16
NA	Sin información	Total	205	139	66	0	446	399	39	8	67	46	9	12
NA	NA	Semana 1	3	2	1	0	10	10	0	0	0	0	0	0
NA	NA	Semana 2	2	2	0	0	6	6	0	0	3	3	0	0
NA	NA	Semana 3	2	2	0	0	2	2	0	0	2	2	0	0

Delimitando el rango de celdas a importar, logré importar los datos de manera adecuada para su posterior limpieza y ordenamiento como se muestra a continuación:

...1	...2	...3	...4	...5	...6	...7	...8	...9	...10	...11	...12	...13	...14	...15
semana 01	2014-12-29	2015-01-04	2016-01-04	2016-01-10	2017-01-02	2017-01-08	2018-01-01	2018-01-07	2018-12-31	2019-01-06	2019-12-30	2020-01-05	2021-01-04	2021-01-10
Semana 02	2015-01-05	2015-01-11	2016-01-11	2016-01-17	2017-01-09	2017-01-15	2018-01-08	2018-01-14	2019-01-07	2019-01-13	2020-01-06	2020-01-12	2021-01-11	2021-01-17
Semana 03	2015-01-12	2015-01-18	2016-01-18	2016-01-24	2017-01-16	2017-01-22	2018-01-15	2018-01-21	2019-01-14	2019-01-20	2020-01-13	2020-01-19	2021-01-18	2021-01-24

...	1	...	2	...	3	...	4	...	5	...	6	...	7	...	8	...	9	...	10	...	11	...	12	...	13	...	14	...	15
Semana	2015-04	2015-01-19	2015-01-25	2016-01-01	2016-01-01	2016-01-01	2016-01-01	2017-01-01	2017-01-01	2017-01-01	2017-01-01	2017-01-01	2017-01-01	2018-01-01	2018-01-01	2018-01-01	2018-01-01	2019-01-01	2019-01-01	2019-01-01	2019-01-01	2020-01-01	2020-01-01	2020-01-01	2020-01-01	2021-01-01	2021-01-01		
Semana	2015-05	2015-01-26	2015-02-01	2016-02-01	2016-02-01	2016-02-01	2016-02-01	2017-02-01	2017-02-01	2017-02-01	2017-02-01	2017-02-01	2017-02-01	2018-02-01	2018-02-01	2018-02-01	2018-02-01	2019-02-01	2019-02-01	2019-02-01	2019-02-01	2020-02-01	2020-02-01	2020-02-01	2020-02-01	2021-02-01	2021-02-01		
Semana	2015-06	2015-02-02	2015-02-08	2016-02-08	2016-02-08	2016-02-08	2016-02-08	2017-02-08	2017-02-08	2017-02-08	2017-02-08	2017-02-08	2017-02-08	2018-02-08	2018-02-08	2018-02-08	2018-02-08	2019-02-08	2019-02-08	2019-02-08	2019-02-08	2020-02-08	2020-02-08	2020-02-08	2020-02-08	2021-02-08	2021-02-08		
Semana	2015-07	2015-02-09	2015-02-15	2016-02-15	2016-02-15	2016-02-15	2016-02-15	2017-02-15	2017-02-15	2017-02-15	2017-02-15	2017-02-15	2017-02-15	2018-02-15	2018-02-15	2018-02-15	2018-02-15	2019-02-15	2019-02-15	2019-02-15	2019-02-15	2020-02-15	2020-02-15	2020-02-15	2020-02-15	2021-02-15	2021-02-15		
Semana	2015-08	2015-02-16	2015-02-22	2016-02-22	2016-02-22	2016-02-22	2016-02-22	2017-02-22	2017-02-22	2017-02-22	2017-02-22	2017-02-22	2017-02-22	2018-02-22	2018-02-22	2018-02-22	2018-02-22	2019-02-22	2019-02-22	2019-02-22	2019-02-22	2020-02-22	2020-02-22	2020-02-22	2020-02-22	2021-02-22	2021-02-22		
Semana	2015-09	2015-02-23	2015-03-01	2016-03-01	2016-03-01	2016-03-01	2016-03-01	2017-03-01	2017-03-01	2017-03-01	2017-03-01	2017-03-01	2017-03-01	2018-03-01	2018-03-01	2018-03-01	2018-03-01	2019-03-01	2019-03-01	2019-03-01	2019-03-01	2020-03-01	2020-03-01	2020-03-01	2020-03-01	2021-03-01	2021-03-01		
Semana	2015-10	2015-03-02	2015-03-08	2016-03-08	2016-03-08	2016-03-08	2016-03-08	2017-03-08	2017-03-08	2017-03-08	2017-03-08	2017-03-08	2017-03-08	2018-03-08	2018-03-08	2018-03-08	2018-03-08	2019-03-08	2019-03-08	2019-03-08	2019-03-08	2020-03-08	2020-03-08	2020-03-08	2020-03-08	2021-03-08	2021-03-08		

...	1	...	2	...	3	...	4	...	5	...	6	...	7	...	8	...	9	...	10	...	11	...	12	...	13	...	14	...	15
2015 Total			NA	193256	98341	94915	0	28803	24331	4446	26	1435	979	440	16														
NA Sin			Total	205	139	66	0	446	399	39	8	67	46	9	12														
informa- ción																													
NA NA			Semana 1	3	2	1	0	10	10	0	0	0	0	0	0														
NA NA			Semana 2	2	2	0	0	6	6	0	0	3	3	0	0														
NA NA			Semana 3	2	2	0	0	2	2	0	0	2	2	0	0														
NA NA			Semana 4	1	1	0	0	4	4	0	0	1	0	0	1														
NA NA			Semana 5	1	1	0	0	4	4	0	0	1	1	0	0														
NA NA			Semana 6	1	1	0	0	6	6	0	0	1	1	0	0														

...1	...2	...3	...4	...5	...6	...7	...8	...9	...10	...11	...12	...13	...14	...15
NA	NA	Semana 7	1	1	0	0	7	6	1	0	1	1	0	0
NA	NA	Semana 8	0	0	0	0	7	7	0	0	0	0	0	0

2 Después de la limpieza

Después de todo el proceso de limpieza y ordenamiento, obtuve dos data frames que contienen la información de las semanas y las defunciones respectivamente, y que cumplen con los principios de “tidy data” (cada variable forma una columna y cada observación una fila)

año	semana	inicio	fin
2015	1	2014-12-29	2015-01-04
2015	2	2015-01-05	2015-01-11
2015	3	2015-01-12	2015-01-18
2015	4	2015-01-19	2015-01-25
2015	5	2015-01-26	2015-02-01
2015	6	2015-02-02	2015-02-08
2015	7	2015-02-09	2015-02-15
2015	8	2015-02-16	2015-02-22
2015	9	2015-02-23	2015-03-01
2015	10	2015-03-02	2015-03-08

año	semana	departamento	sexo	natural	violenta	en estudio
2015	1	Amazonas	masculino	4	2	0
2015	1	Amazonas	femenino	2	0	0
2015	1	Amazonas	indeterminado	0	0	0
2015	1	Antioquia	masculino	294	89	4
2015	1	Antioquia	femenino	321	16	2
2015	1	Antioquia	indeterminado	0	0	0
2015	1	Arauca	masculino	7	6	0
2015	1	Arauca	femenino	8	1	0
2015	1	Arauca	indeterminado	0	0	0
2015	1	Atlántico	masculino	107	22	0