

Analysis of Death Cases in United States of America

Sushant Deepak Parkhi

About Dataset

1. It has a data about Multiple Causes of Death in 2014
2. Original Dataset is by **Centers of Disease Control and Prevention**
3. It was released to Public in **December, 2015**
4. There are total of **2,631,171** observations of **38** variables

Tables	Description
Death Record	Primary Table containing a single row per death record
Place of Injury	Look Up Table Related to Place of Injury
Manner of Death	Look Up Table Related to Manner of Death
Education2003Revision	Look Up Table Related to Type of Education
Icd10Code	Look Up Table Related to the Cause of Death

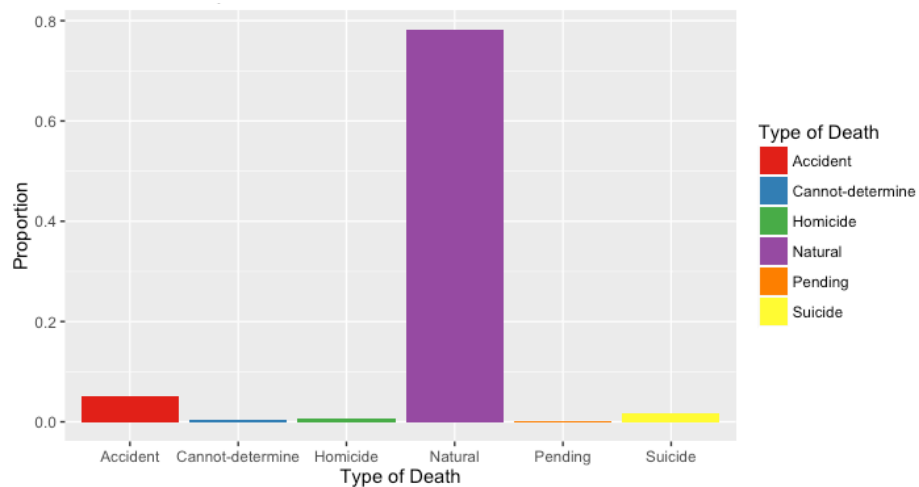
Death Record Table Attributes	Description
Manner of Death	Integer Data Containing Manner of Death
Marital Status	Categorical Data Containing Details About Marital Status
Education2003Revision	Categorical Data Containing Details About Education Status
Icd10Code	Categorical Data Containing Details About the Cause of Death
Sex	Categorical Data Containing Details About the Sex
PlaceOfInjury	Categorical Data Containing Details About the Place of Injury

Dataset Figures

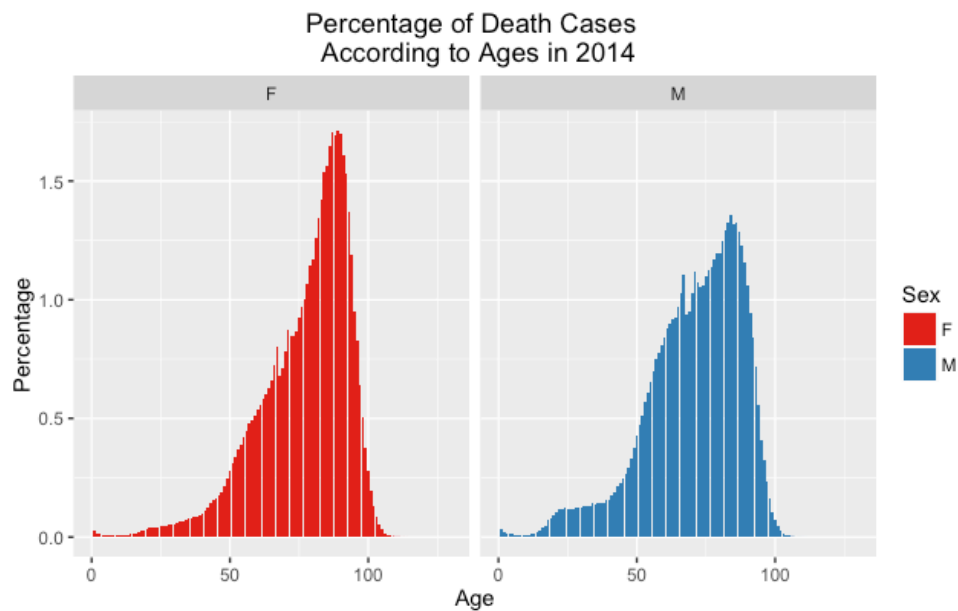
1. There were total of 2,631,171 Death Cases in 2014
2. No. of Female Death Cases were 1,299,710 and No. of Male Death Cases were 1,331,461
3. Most Cause of Death Were Natural, Suicidal and Accidental

Analysis of Numerical and Categorical Data

Proportion of Total Deaths in United States on Manner of Death

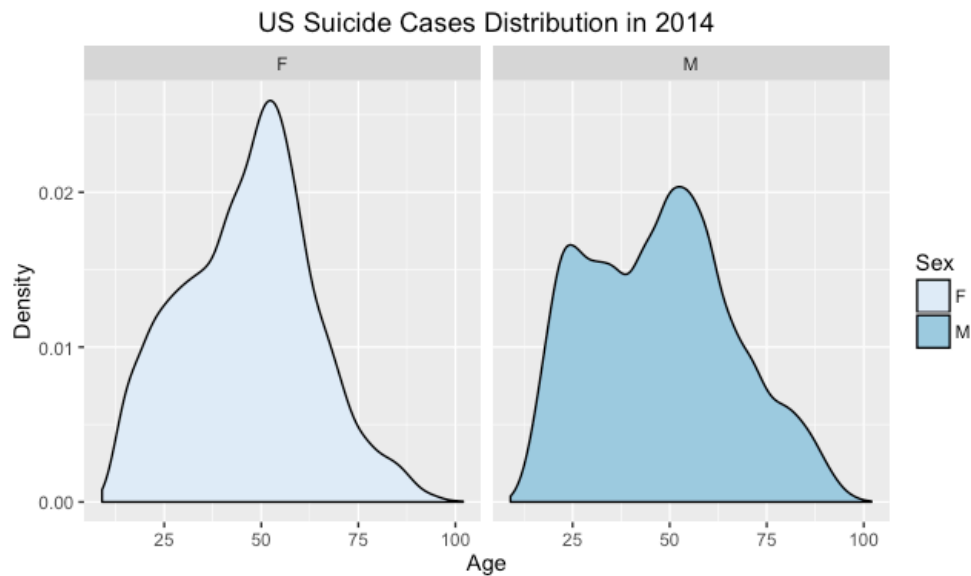


Percentage of Death Cases According to Ages



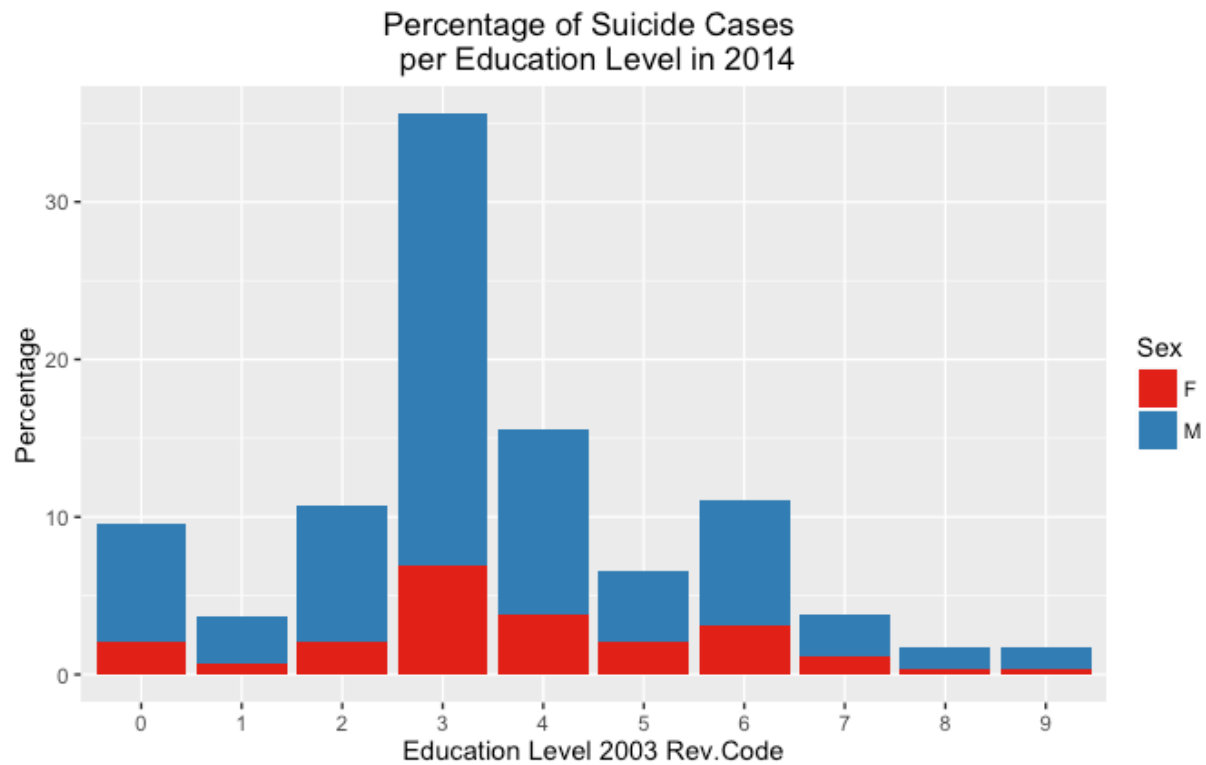
Suicide Cases

	Sex	Cases	Percentage	Mean	Std
	(chr)	(int)	(dbl)	(dbl)	(dbl)
1	F	9775	22.66299	46.57483	16.78307
2	M	33357	77.33701	47.65932	18.85484



By Education

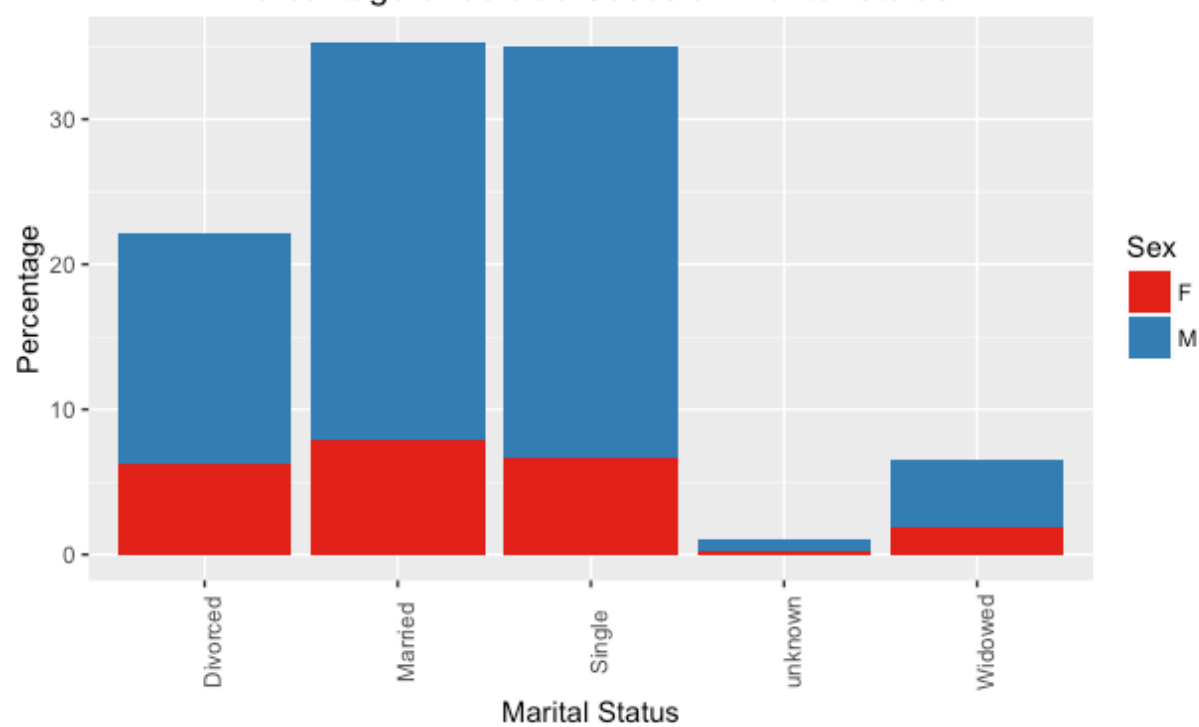
	Education2003Revision	Sex	Sum	Percentage
	(int)	(chr)	(int)	(dbl)
1	0	F	903	2.0935732
2	0	M	3213	7.4492256
3	1	F	299	0.6932208
4	1	M	1278	2.9629973
5	2	F	917	2.1260317
6	2	M	3705	8.5899100
7	3	F	2995	6.9438004
8	3	M	12361	28.6585366
9	4	F	1661	3.8509691
10	4	M	5065	11.7430214
11	5	F	872	2.0217008
12	5	M	1985	4.6021515
13	6	F	1361	3.1554298
14	6	M	3393	7.8665492
15	7	F	475	1.1012705
16	7	M	1143	2.6500046
17	8	F	125	0.2898080
18	8	M	610	1.4142632
19	9	F	167	0.3871835
20	9	M	604	1.4003524



By Marital Status

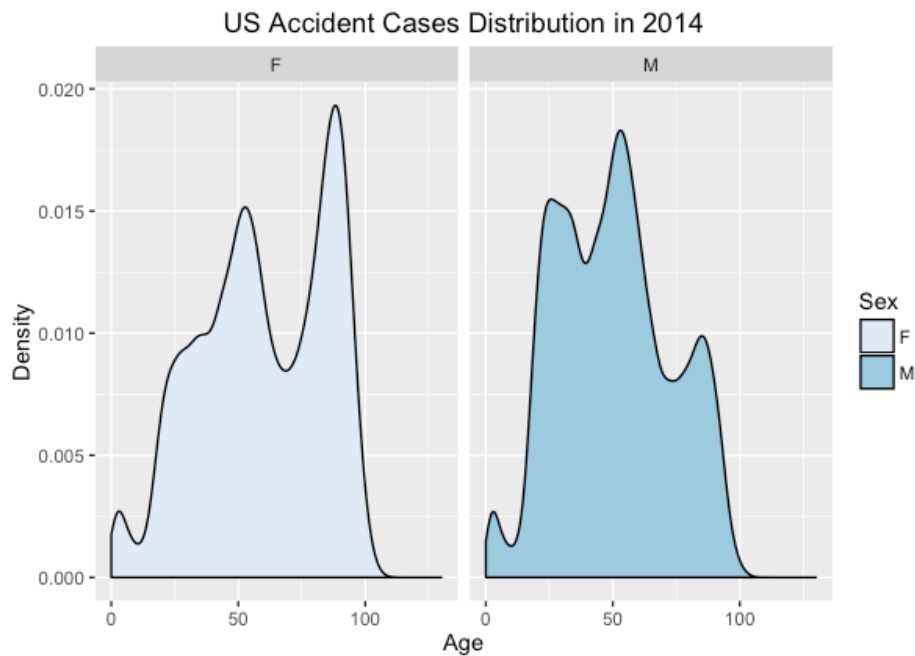
	MaritalStatus (chr)	Sex (chr)	Cases_Suicide (int)	Percentage_Suicide (dbl)
1	D	F	2676	6.2042103
2	D	M	6881	15.9533525
3	M	F	3380	7.8364092
4	M	M	11867	27.5132152
5	S	F	2849	6.6053046
6	S	M	12247	28.3942317
7	U	F	81	0.1877956
8	U	M	353	0.8184179
9	W	F	789	1.8292683
10	W	M	2009	4.6577947

Percentage of Suicide Cases on Marital Status



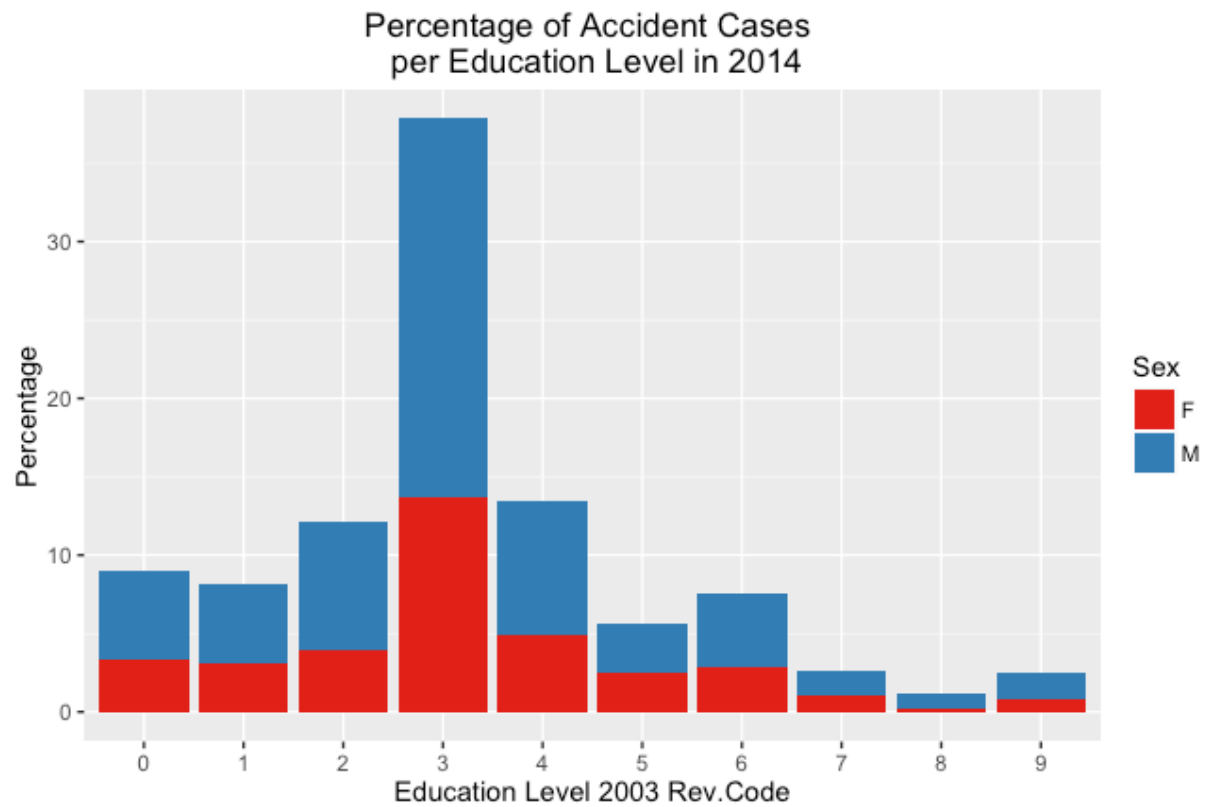
Accident Cases

	Sex (chr)	Cases (int)	Percentage (dbl)	Mean (dbl)	Std (dbl)
1	F	48352	36.45411	59.87610	25.72957
2	M	84286	63.54589	50.48407	23.91828



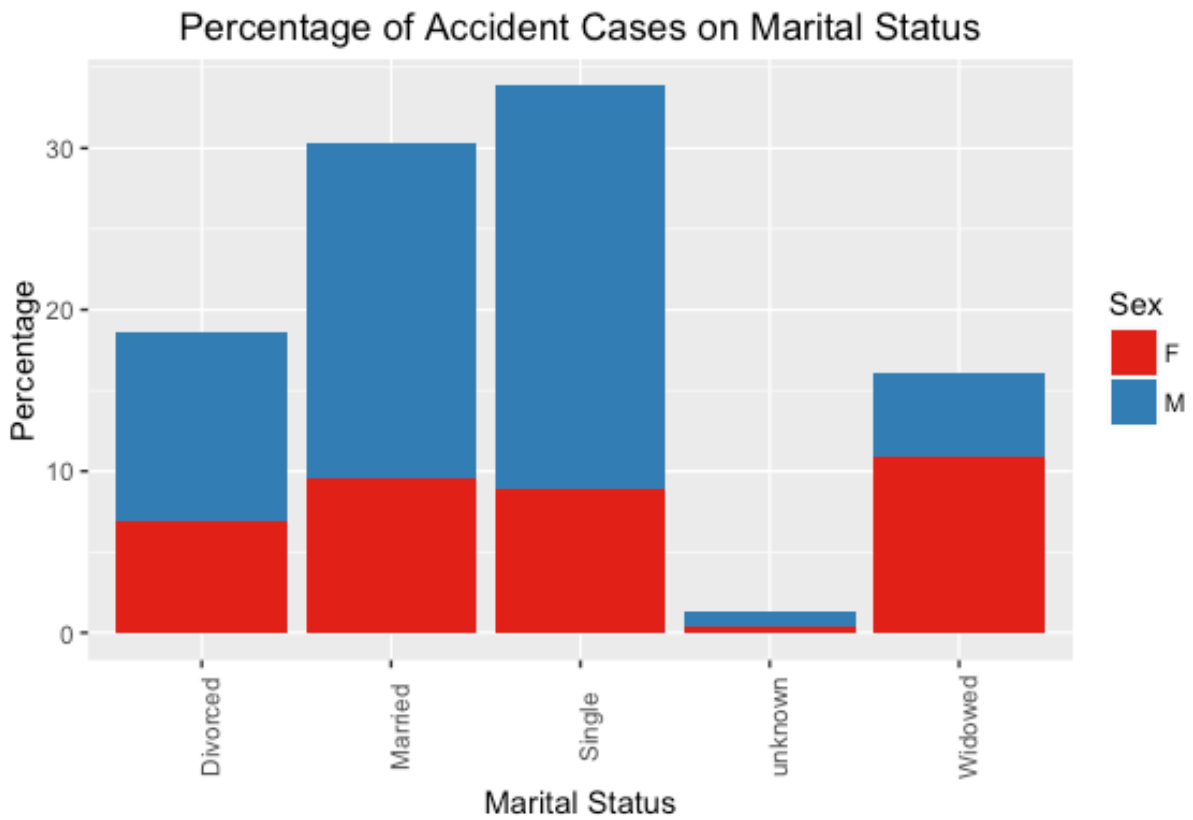
By Education

	Education2003Revision (int)	Sex (chr)	Sum (int)	Percentage (dbl)
1	0	F	4419	3.3316244
2	0	M	7461	5.6250848
3	1	F	4153	3.1310786
4	1	M	6710	5.0588821
5	2	F	5295	3.9920686
6	2	M	10713	8.0768709
7	3	F	18161	13.6921546
8	3	M	32032	24.1499419
9	4	F	6582	4.9623788
10	4	M	11251	8.4824862
11	5	F	3290	2.4804355
12	5	M	4149	3.1280628
13	6	F	3771	2.8430766
14	6	M	6326	4.7693723
15	7	F	1366	1.0298708
16	7	M	2121	1.5990893
17	8	F	302	0.2276874
18	8	M	1223	0.9220585
19	9	F	1013	0.7637329
20	9	M	2300	1.7340430



By Marital Status

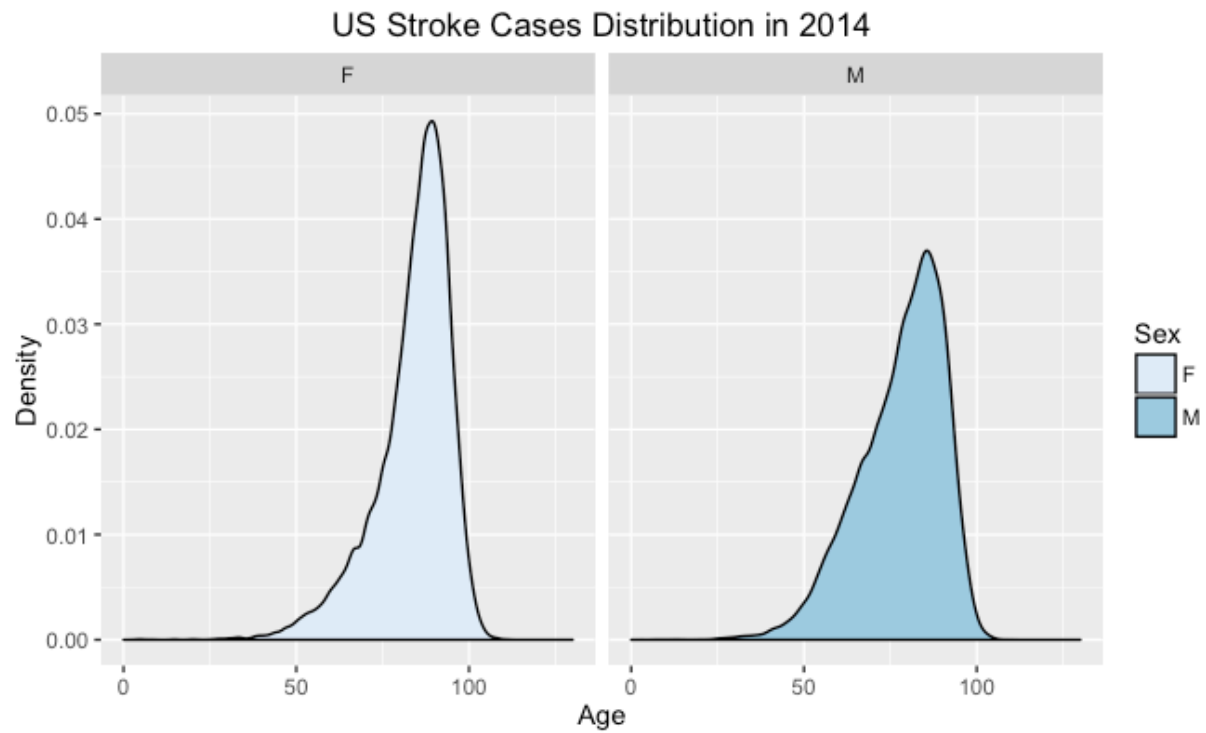
	MaritalStatus (chr)	Sex (chr)	Cases_Suicide (int)	Percentage_Accident (dbl)
1	D	F	9147	6.8962138
2	D	M	15476	11.6678478
3	M	F	12614	9.5100951
4	M	M	27559	20.7776052
5	S	F	11758	8.8647296
6	S	M	33141	24.9860523
7	U	F	419	0.3158974
8	U	M	1272	0.9590012
9	W	F	14414	10.8671723
10	W	M	6838	5.1553853



Stroke Cases

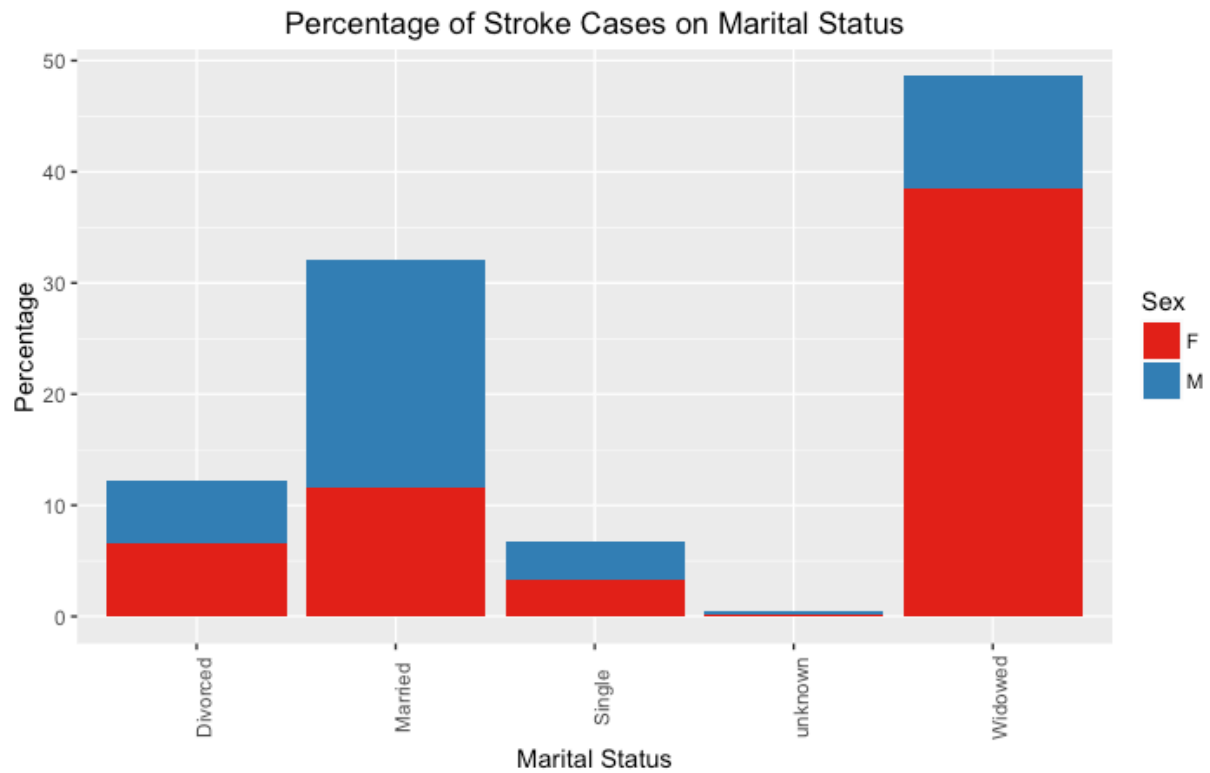
Filtering Data by ICD10Code

	Sex	Cases	Percentage	Mean	Std
	(chr)	(int)	(dbl)	(dbl)	(dbl)
1	F	39465	60.18391	83.71131	11.04510
2	M	26109	39.81609	77.98885	12.33413



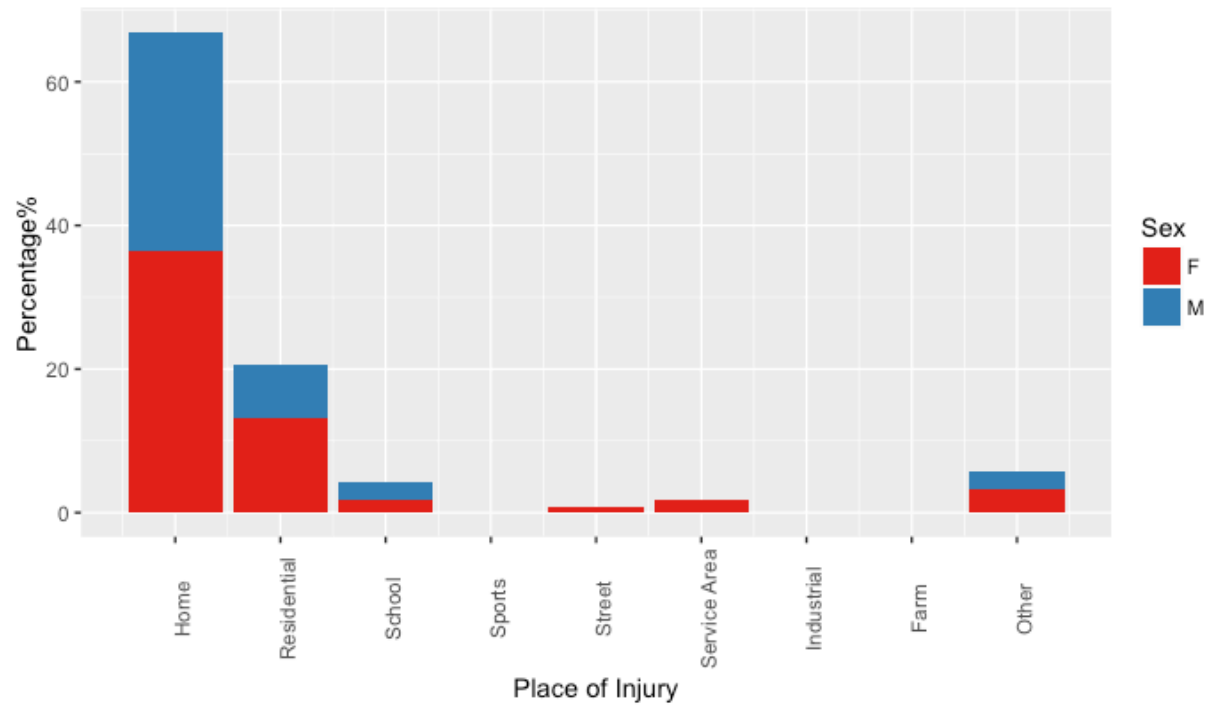
By Marital Status

	MaritalStatus	Sex	Cases_	Percentage_Stroke
	(chr)	(chr)	(int)	(dbl)
1	D	F	4346	6.6276268
2	D	M	3630	5.5357306
3	M	F	7572	11.5472596
4	M	M	13485	20.5645530
5	S	F	2176	3.3183884
6	S	M	2191	3.3412633
7	U	F	96	0.1463995
8	U	M	191	0.2912740
9	W	F	25275	38.5442401
10	W	M	6612	10.0832647



By Place of Injury

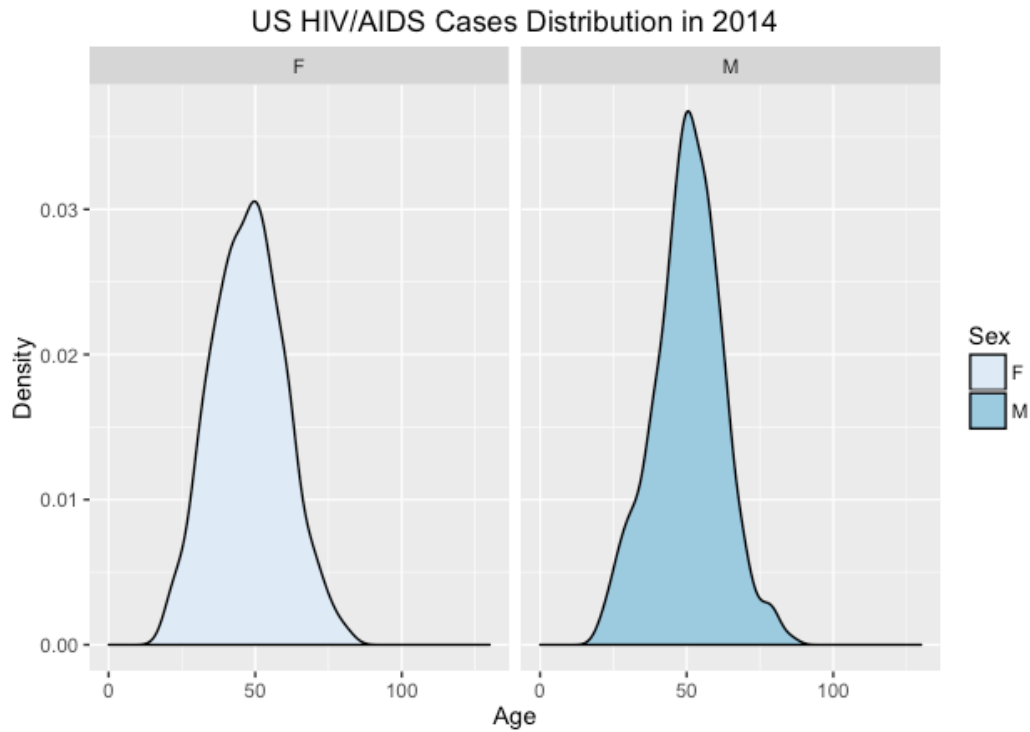
	PlaceOfInjury	Sex	Cases_Strokes	Percentage_Stokes
	(int)	(chr)	(int)	(dbl)
1	0	F	44	36.3636364
2	0	M	37	30.5785124
3	1	F	16	13.2231405
4	1	M	9	7.4380165
5	2	F	2	1.6528926
6	2	M	3	2.4793388
7	4	F	1	0.8264463
8	5	F	2	1.6528926
9	8	F	4	3.3057851
10	8	M	3	2.4793388



HIV/AIDS Cases

Filtered Dataset by ICD10Code

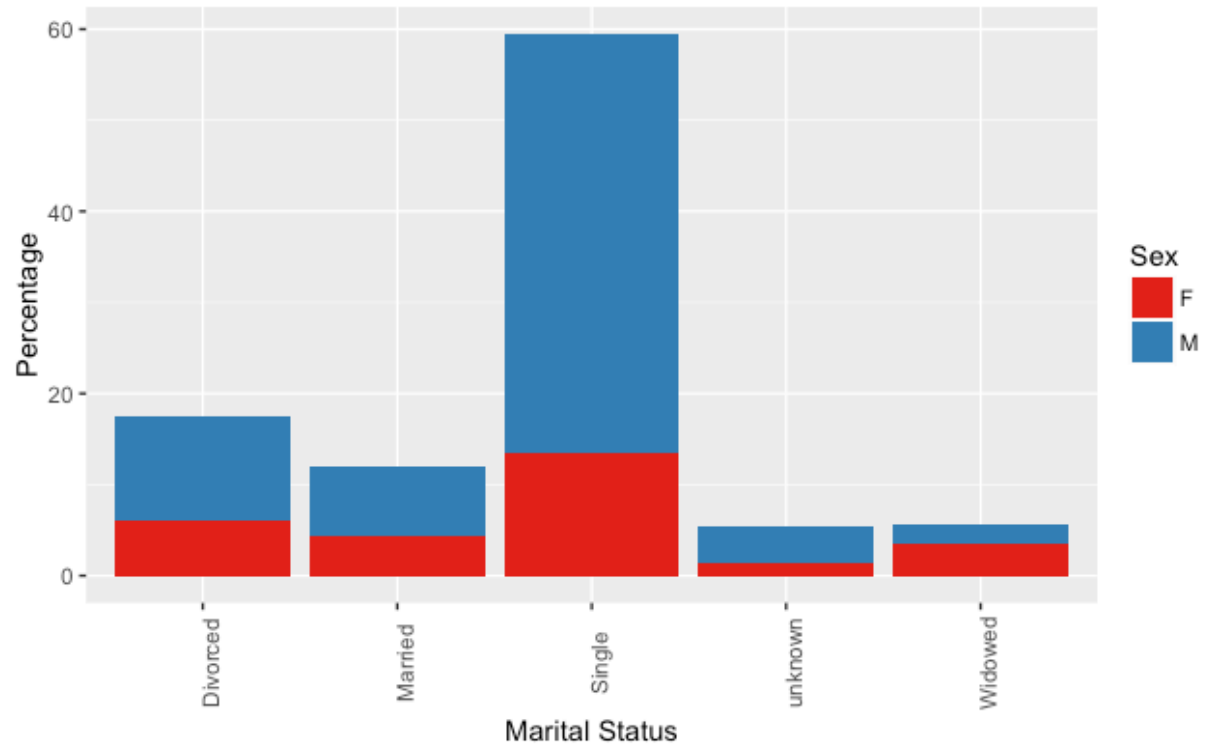
	Sex	Cases	Percentage	Mean	Std
	(chr)	(int)	(dbl)	(dbl)	(dbl)
1	F	417	28.89813	47.70024	12.18818
2	M	1026	71.10187	50.53314	11.67434



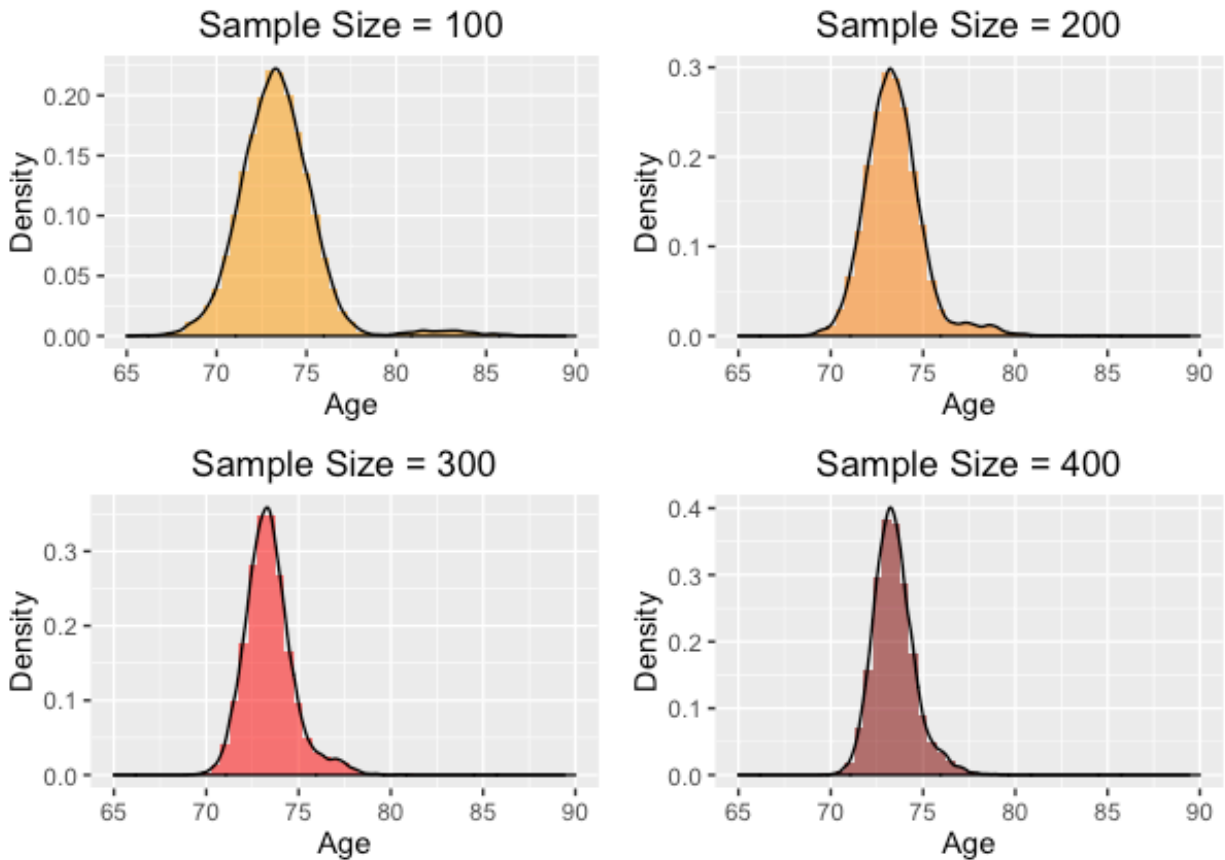
By Marital Status

	MaritalStatus	Sex	Cases_	Percentage_Aids
	(chr)	(chr)	(int)	(dbl)
1	D	F	88	6.098406
2	D	M	166	11.503812
3	M	F	62	4.296604
4	M	M	111	7.692308
5	S	F	195	13.513514
6	S	M	664	46.015246
7	U	F	20	1.386001
8	U	M	57	3.950104
9	W	F	52	3.603604
10	W	M	28	1.940402

Percentage of HIV/AIDS Cases on Marital Status



Central Limit Theorem



```
> cat("(Sample Size = 100) Mean = ",mean(xbar_100),"Standard Deviation = ",sd(xbar_100),"\\n")
(Sample Size = 100) Mean = 73.39729 Standard Deviation = 2.241673
> cat("(Sample Size = 200) Mean = ",mean(xbar_200),"Standard Deviation = ",sd(xbar_200),"\\n")
(Sample Size = 200) Mean = 73.42346 Standard Deviation = 1.600213
> cat("(Sample Size = 300) Mean = ",mean(xbar_300),"Standard Deviation = ",sd(xbar_300),"\\n")
(Sample Size = 300) Mean = 73.40492 Standard Deviation = 1.321567
> cat("(Sample Size = 400) Mean = ",mean(xbar_400),"Standard Deviation = ",sd(xbar_400),"\\n")
(Sample Size = 400) Mean = 73.42934 Standard Deviation = 1.133104
```

Sampling

Simple Sampling

	Manner of Death	Frequency	Description
1	0	15	Non Specified
2	1	6	Accident
3	3	1	Homicide
4	4	1	Pending Investigation
5	7	77	Natural

Systematic Sampling

	Manner of Death	Frequency	Description
1	0	364034	Non Specified
2	1	132684	Accident
3	2	43139	Suicide
4	3	16840	Homicide
5	4	3703	Pending Investigation
6	5	10838	Could not determine
7	7	2059933	Natural

Confidence Interval

Simple Sampling

80% Conf Level ($\alpha = 0.20$), CI = 73.40 - 73.43

90% Conf Level ($\alpha = 0.10$), CI = 73.39 - 73.44

Systematic Sampling

80% Conf Level ($\alpha = 0.20$), CI = 73.40 - 73.43

90% Conf Level ($\alpha = 0.10$), CI = 73.39 - 73.44