

# LA-report

Praveen, Rishina

7/12/2021

## Contents

<b>1</b>	<b>EXPLORATORY DATA ANALYSIS</b>	<b>1</b>
<b>2</b>	<b>PLOTS</b>	<b>13</b>
2.1	Histograms . . . . .	13
2.2	GGplots . . . . .	16

## 1 EXPLORATORY DATA ANALYSIS

We start with importing our data set and storing it in data frame “data”, using read.csv function.

```
data <- read.csv("C:\\Users\\LENOVO\\OneDrive\\Desktop\\praveen\\EDA\\LA-1\\USA Covid Data.csv", header = TRUE)
show(data)
```

##	State	Total.Cases	Total.Deaths	Total.Recovered	Active.Cases
## 1	Alabama	551298	11358	NA	NA
## 2	Alaska	68674	370	66087	2217
## 3	Arizona	898283	18004	867042	13237
## 4	Arkansas	353095	5933	341896	5266
## 5	California	3828483	63758	2077558	1687167
## 6	Colorado	560419	7056	467669	85694
## 7	Connecticut	349743	8278	339469	1996
## 8	Delaware	109900	1695	107493	712
## 9	Florida	2382980	37985	2215966	129029
## 10	Georgia	1137560	21467	1056403	59690
## 11	Hawaii	38121	518	NA	NA
## 12	Idaho	195552	2162	115460	77930
## 13	Illinois	1394235	25733	1358187	10315
## 14	Indiana	756070	13892	730685	11493
## 15	Iowa	405303	6149	366566	32588
## 16	Kansas	319586	5166	309373	5047
## 17	Kentucky	466244	7250	53643	405351
## 18	Louisiana	484577	10763	NA	NA
## 19	Maine	69156	861	NA	NA
## 20	Maryland	462736	9755	11041	441940

## 21	Massachusetts	710519	18008	691163	1348
## 22	Michigan	1000908	21031	868294	111583
## 23	Minnesota	606034	7693	597632	709
## 24	Mississippi	323214	7436	312956	2822
## 25	Missouri	626126	10114	562379	53633
## 26	Montana	114102	1671	112003	428
## 27	Nebraska	224873	2262	NA	NA
## 28	Nevada	336562	5705	315944	14913
## 29	New Hampshire	99652	1374	98083	195
## 30	New Jersey	1024748	26490	983195	15063
## 31	New Mexico	205996	4347	195187	6462
## 32	New York	2176123	54065	1992360	129698
## 33	North Carolina	1016262	13460	994472	8330
## 34	North Dakota	110779	1528	109127	124
## 35	Ohio	1113006	20366	1085985	6655
## 36	Oklahoma	459675	7406	449960	2309
## 37	Oregon	209764	2788	195932	11044
## 38	Pennsylvania	1217879	27821	1176921	13137
## 39	Puerto Rico	275539	2552	121227	151760
## 40	South Carolina	598229	9843	575173	13213
## 41	South Dakota	124641	2039	122424	178
## 42	Tennessee	868738	12583	853563	2592
## 43	Texas	3010445	52722	2903306	54417
## 44	Utah	417653	2387	407664	7602
## 45	Vermont	24440	258	24067	115
## 46	Virginia	681599	11436	57321	612842
## 47	Washington	456423	6029	227194	223200
## 48	West Virginia	164346	2901	160390	1055
## 49	Wisconsin	678232	8154	NA	NA
## 50	Wyoming	62737	751	61420	566
##	Total.Cases.1.mil.population		Death.1.mil.population	Total.Tests	
## 1		112437		2316	2830286
## 2		93875		506	2395221
## 3		123412		2474	5126795
## 4		117004		1966	3808827
## 5		96894		1614	70392011
## 6		97316		1225	3190081
## 7		98097		2322	9645449
## 8		112861		1741	740119
## 9		110951		1769	31249599
## 10		107141		2022	12006950
## 11		26924		366	1754418
## 12		109426		1210	1413139
## 13		110026		2031	25989308
## 14		112306		2064	10918291
## 15		128461		1949	5216309
## 16		109698		1773	1455337
## 17		104359		1623	6941164
## 18		104237		2315	7810833
## 19		51447		641	2765484
## 20		76540		1614	10927231
## 21		103086		2613	23969707
## 22		100223		2106	14953233
## 23		107460		1364	10377218

## 24	108601	2499	2911438
## 25	102018	1648	7800545
## 26	106759	1563	1457325
## 27	116249	1169	3056982
## 28	109268	1852	3555730
## 29	73289	1011	2448734
## 30	115371	2982	14458068
## 31	98242	2073	3686555
## 32	111862	2779	59267467
## 33	96897	1283	13902865
## 34	145367	2005	447733
## 35	95217	1742	13728606
## 36	116168	1872	4040035
## 37	49734	661	5471841
## 38	95132	2173	14680614
## 39	81353	753	464073
## 40	116190	1912	8227548
## 41	140892	2305	490627
## 42	127210	1843	8227135
## 43	103823	1818	32192208
## 44	130274	745	5141040
## 45	39167	413	1741050
## 46	79854	1340	10293262
## 47	59938	792	7673465
## 48	91703	1619	3030600
## 49	116486	1400	3604344
## 50	108399	1298	791749
##	Tests.1.mil.population	Population	
## 1	577234	4903185	
## 2	3274195	731545	
## 3	704354	7278717	
## 4	1262119	3017804	
## 5	1781525	39512223	
## 6	553955	5758736	
## 7	2705378	3565287	
## 8	760060	973764	
## 9	1454976	21477737	
## 10	1130872	10617423	
## 11	1239108	1415872	
## 12	790760	1787065	
## 13	2050953	12671821	
## 14	1621797	6732219	
## 15	1653310	3155070	
## 16	499547	2913314	
## 17	1553642	4467673	
## 18	1680185	4648794	
## 19	2057327	1344212	
## 20	1807444	6045680	
## 21	3477649	6892503	
## 22	1497291	9986857	
## 23	1840052	5639632	
## 24	978257	2976149	
## 25	1270979	6137428	
## 26	1363543	1068778	

```
## 27      1580319      1934408
## 28      1154399      3080156
## 29      1800922      1359711
## 30      1627759      8882190
## 31      1758157      2096829
## 32      3046613      19453561
## 33      1325587      10488084
## 34          587528          762062
## 35      1174479      11689100
## 36      1020992      3956971
## 37      1297340      4217737
## 38      1146745      12801989
## 39          137018      3386941
## 40      1597981      5148714
## 41          554594      884659
## 42      1204704      6829174
## 43      1110234      28995881
## 44      1603589      3205958
## 45      2790193      623989
## 46      1205933      8535519
## 47      1007692      7614893
## 48      1691044      1792147
## 49          619044      5822434
## 50      1368012      578759
```

This data set we are working on is state-wise stats for Covid-19 in USA. The data contains the several stats like active cases, total cases, total recovery etc.

to make our data more rich, we will add a few more columns, based on our need.

```
f_data <- mutate(data, Percent_Death = (data$Total.Deaths*100)/data$Total.Cases)
f_data <- mutate(f_data, Percent.Recovered = (data$Total.Recovered*100)/data$Total.Cases)

show(f_data)
```

```
##      State Total.Cases Total.Deaths Total.Recovered Active.Cases
## 1  Alabama      551298          11358              NA           NA
## 2   Alaska      68674           370          66087          2217
## 3   Arizona      898283          18004          867042          13237
## 4   Arkansas      353095          5933          341896          5266
## 5  California     3828483          63758          2077558          1687167
## 6   Colorado      560419          7056          467669          85694
## 7  Connecticut      349743          8278          339469          1996
## 8   Delaware      109900          1695          107493           712
## 9   Florida     2382980          37985          2215966          129029
## 10  Georgia     1137560          21467          1056403          59690
## 11   Hawaii       38121           518              NA           NA
## 12   Idaho      195552          2162          115460          77930
## 13  Illinois     1394235          25733          1358187          10315
## 14   Indiana      756070          13892          730685          11493
## 15    Iowa       405303          6149          366566          32588
## 16   Kansas      319586          5166          309373          5047
## 17  Kentucky      466244          7250          53643          405351
## 18  Louisiana      484577          10763              NA           NA
```

## 19	Maine	69156	861	NA	NA
## 20	Maryland	462736	9755	11041	441940
## 21	Massachusetts	710519	18008	691163	1348
## 22	Michigan	1000908	21031	868294	111583
## 23	Minnesota	606034	7693	597632	709
## 24	Mississippi	323214	7436	312956	2822
## 25	Missouri	626126	10114	562379	53633
## 26	Montana	114102	1671	112003	428
## 27	Nebraska	224873	2262	NA	NA
## 28	Nevada	336562	5705	315944	14913
## 29	New Hampshire	99652	1374	98083	195
## 30	New Jersey	1024748	26490	983195	15063
## 31	New Mexico	205996	4347	195187	6462
## 32	New York	2176123	54065	1992360	129698
## 33	North Carolina	1016262	13460	994472	8330
## 34	North Dakota	110779	1528	109127	124
## 35	Ohio	1113006	20366	1085985	6655
## 36	Oklahoma	459675	7406	449960	2309
## 37	Oregon	209764	2788	195932	11044
## 38	Pennsylvania	1217879	27821	1176921	13137
## 39	Puerto Rico	275539	2552	121227	151760
## 40	South Carolina	598229	9843	575173	13213
## 41	South Dakota	124641	2039	122424	178
## 42	Tennessee	868738	12583	853563	2592
## 43	Texas	3010445	52722	2903306	54417
## 44	Utah	417653	2387	407664	7602
## 45	Vermont	24440	258	24067	115
## 46	Virginia	681599	11436	57321	612842
## 47	Washington	456423	6029	227194	223200
## 48	West Virginia	164346	2901	160390	1055
## 49	Wisconsin	678232	8154	NA	NA
## 50	Wyoming	62737	751	61420	566
##	Total.Cases.1.mil.population	Death.1.mil.population	Total.Tests		
## 1	112437	2316	2830286		
## 2	93875	506	2395221		
## 3	123412	2474	5126795		
## 4	117004	1966	3808827		
## 5	96894	1614	70392011		
## 6	97316	1225	3190081		
## 7	98097	2322	9645449		
## 8	112861	1741	740119		
## 9	110951	1769	31249599		
## 10	107141	2022	12006950		
## 11	26924	366	1754418		
## 12	109426	1210	1413139		
## 13	110026	2031	25989308		
## 14	112306	2064	10918291		
## 15	128461	1949	5216309		
## 16	109698	1773	1455337		
## 17	104359	1623	6941164		
## 18	104237	2315	7810833		
## 19	51447	641	2765484		
## 20	76540	1614	10927231		
## 21	103086	2613	23969707		

## 22	100223	2106	14953233	
## 23	107460	1364	10377218	
## 24	108601	2499	2911438	
## 25	102018	1648	7800545	
## 26	106759	1563	1457325	
## 27	116249	1169	3056982	
## 28	109268	1852	3555730	
## 29	73289	1011	2448734	
## 30	115371	2982	14458068	
## 31	98242	2073	3686555	
## 32	111862	2779	59267467	
## 33	96897	1283	13902865	
## 34	145367	2005	447733	
## 35	95217	1742	13728606	
## 36	116168	1872	4040035	
## 37	49734	661	5471841	
## 38	95132	2173	14680614	
## 39	81353	753	464073	
## 40	116190	1912	8227548	
## 41	140892	2305	490627	
## 42	127210	1843	8227135	
## 43	103823	1818	32192208	
## 44	130274	745	5141040	
## 45	39167	413	1741050	
## 46	79854	1340	10293262	
## 47	59938	792	7673465	
## 48	91703	1619	3030600	
## 49	116486	1400	3604344	
## 50	108399	1298	791749	
##	Tests.1.mil.population	Population	Percent_Death	Percent.Recovered
## 1	577234	4903185	2.0602288	NA
## 2	3274195	731545	0.5387774	96.232927
## 3	704354	7278717	2.0042681	96.522143
## 4	1262119	3017804	1.6802843	96.828332
## 5	1781525	39512223	1.6653594	54.265828
## 6	553955	5758736	1.2590580	83.449883
## 7	2705378	3565287	2.3668808	97.062414
## 8	760060	973764	1.5423112	97.809827
## 9	1454976	21477737	1.5940125	92.991381
## 10	1130872	10617423	1.8871093	92.865695
## 11	1239108	1415872	1.3588311	NA
## 12	790760	1787065	1.1055883	59.043119
## 13	2050953	12671821	1.8456716	97.414496
## 14	1621797	6732219	1.8373960	96.642507
## 15	1653310	3155070	1.5171366	90.442459
## 16	499547	2913314	1.6164663	96.804303
## 17	1553642	4467673	1.5549798	11.505349
## 18	1680185	4648794	2.2211124	NA
## 19	2057327	1344212	1.2450113	NA
## 20	1807444	6045680	2.1081135	2.386026
## 21	3477649	6892503	2.5344854	97.275794
## 22	1497291	9986857	2.1011921	86.750630
## 23	1840052	5639632	1.2694007	98.613609
## 24	978257	2976149	2.3006429	96.826251

## 25	1270979	6137428	1.6153298	89.818822
## 26	1363543	1068778	1.4644792	98.160418
## 27	1580319	1934408	1.0059011	NA
## 28	1154399	3080156	1.6950814	93.873937
## 29	1800922	1359711	1.3787982	98.425521
## 30	1627759	8882190	2.5850258	95.945052
## 31	1758157	2096829	2.1102352	94.752811
## 32	3046613	19453561	2.4844643	91.555487
## 33	1325587	10488084	1.3244616	97.855868
## 34	587528	762062	1.3793228	98.508743
## 35	1174479	11689100	1.8298194	97.572250
## 36	1020992	3956971	1.6111383	97.886550
## 37	1297340	4217737	1.3291127	93.405923
## 38	1146745	12801989	2.2843813	96.636940
## 39	137018	3386941	0.9261847	43.996313
## 40	1597981	5148714	1.6453565	96.145957
## 41	554594	884659	1.6358983	98.221292
## 42	1204704	6829174	1.4484229	98.253213
## 43	1110234	28995881	1.7513025	96.441091
## 44	1603589	3205958	0.5715271	97.608302
## 45	2790193	623989	1.0556465	98.473813
## 46	1205933	8535519	1.6778194	8.409783
## 47	1007692	7614893	1.3209238	49.777071
## 48	1691044	1792147	1.7651783	97.592883
## 49	619044	5822434	1.2022435	NA
## 50	1368012	578759	1.1970607	97.900760

Now, we will add values at places that have no values, except in the 2 columns that we have manually added.  
We add the mean of that column.

```
f_data$Total.Recovered[is.na(f_data$Total.Recovered)]<- mean(f_data$Total.Recovered,na.rm=TRUE)
f_data$Active.Cases[is.na(f_data$Active.Cases)]<- mean(f_data$Active.Cases,na.rm=TRUE)
show(f_data)
```

##	State	Total.Cases	Total.Deaths	Total.Recovered	Active.Cases
## 1	Alabama	551298	11358	607724.5	100356
## 2	Alaska	68674	370	66087.0	2217
## 3	Arizona	898283	18004	867042.0	13237
## 4	Arkansas	353095	5933	341896.0	5266
## 5	California	3828483	63758	2077558.0	1687167
## 6	Colorado	560419	7056	467669.0	85694
## 7	Connecticut	349743	8278	339469.0	1996
## 8	Delaware	109900	1695	107493.0	712
## 9	Florida	2382980	37985	2215966.0	129029
## 10	Georgia	1137560	21467	1056403.0	59690
## 11	Hawaii	38121	518	607724.5	100356
## 12	Idaho	195552	2162	115460.0	77930
## 13	Illinois	1394235	25733	1358187.0	10315
## 14	Indiana	756070	13892	730685.0	11493
## 15	Iowa	405303	6149	366566.0	32588
## 16	Kansas	319586	5166	309373.0	5047

## 17	Kentucky	466244	7250	53643.0	405351
## 18	Louisiana	484577	10763	607724.5	100356
## 19	Maine	69156	861	607724.5	100356
## 20	Maryland	462736	9755	11041.0	441940
## 21	Massachusetts	710519	18008	691163.0	1348
## 22	Michigan	1000908	21031	868294.0	111583
## 23	Minnesota	606034	7693	597632.0	709
## 24	Mississippi	323214	7436	312956.0	2822
## 25	Missouri	626126	10114	562379.0	53633
## 26	Montana	114102	1671	112003.0	428
## 27	Nebraska	224873	2262	607724.5	100356
## 28	Nevada	336562	5705	315944.0	14913
## 29	New Hampshire	99652	1374	98083.0	195
## 30	New Jersey	1024748	26490	983195.0	15063
## 31	New Mexico	205996	4347	195187.0	6462
## 32	New York	2176123	54065	1992360.0	129698
## 33	North Carolina	1016262	13460	994472.0	8330
## 34	North Dakota	110779	1528	109127.0	124
## 35	Ohio	1113006	20366	1085985.0	6655
## 36	Oklahoma	459675	7406	449960.0	2309
## 37	Oregon	209764	2788	195932.0	11044
## 38	Pennsylvania	1217879	27821	1176921.0	13137
## 39	Puerto Rico	275539	2552	121227.0	151760
## 40	South Carolina	598229	9843	575173.0	13213
## 41	South Dakota	124641	2039	122424.0	178
## 42	Tennessee	868738	12583	853563.0	2592
## 43	Texas	3010445	52722	2903306.0	54417
## 44	Utah	417653	2387	407664.0	7602
## 45	Vermont	24440	258	24067.0	115
## 46	Virginia	681599	11436	57321.0	612842
## 47	Washington	456423	6029	227194.0	223200
## 48	West Virginia	164346	2901	160390.0	1055
## 49	Wisconsin	678232	8154	607724.5	100356
## 50	Wyoming	62737	751	61420.0	566
##	Total.Cases.1.mil.population		Death.1.mil.population	Total.Tests	
## 1		112437		2316	2830286
## 2		93875		506	2395221
## 3		123412		2474	5126795
## 4		117004		1966	3808827
## 5		96894		1614	70392011
## 6		97316		1225	3190081
## 7		98097		2322	9645449
## 8		112861		1741	740119
## 9		110951		1769	31249599
## 10		107141		2022	12006950
## 11		26924		366	1754418
## 12		109426		1210	1413139
## 13		110026		2031	25989308
## 14		112306		2064	10918291
## 15		128461		1949	5216309
## 16		109698		1773	1455337
## 17		104359		1623	6941164
## 18		104237		2315	7810833
## 19		51447		641	2765484



## 20	76540	1614	10927231	
## 21	103086	2613	23969707	
## 22	100223	2106	14953233	
## 23	107460	1364	10377218	
## 24	108601	2499	2911438	
## 25	102018	1648	7800545	
## 26	106759	1563	1457325	
## 27	116249	1169	3056982	
## 28	109268	1852	3555730	
## 29	73289	1011	2448734	
## 30	115371	2982	14458068	
## 31	98242	2073	3686555	
## 32	111862	2779	59267467	
## 33	96897	1283	13902865	
## 34	145367	2005	447733	
## 35	95217	1742	13728606	
## 36	116168	1872	4040035	
## 37	49734	661	5471841	
## 38	95132	2173	14680614	
## 39	81353	753	464073	
## 40	116190	1912	8227548	
## 41	140892	2305	490627	
## 42	127210	1843	8227135	
## 43	103823	1818	32192208	
## 44	130274	745	5141040	
## 45	39167	413	1741050	
## 46	79854	1340	10293262	
## 47	59938	792	7673465	
## 48	91703	1619	3030600	
## 49	116486	1400	3604344	
## 50	108399	1298	791749	
##	Tests.1.mil.population	Population	Percent_Death	Percent.Recovered
## 1	577234	4903185	2.0602288	NA
## 2	3274195	731545	0.5387774	96.232927
## 3	704354	7278717	2.0042681	96.522143
## 4	1262119	3017804	1.6802843	96.828332
## 5	1781525	39512223	1.6653594	54.265828
## 6	553955	5758736	1.2590580	83.449883
## 7	2705378	3565287	2.3668808	97.062414
## 8	760060	973764	1.5423112	97.809827
## 9	1454976	21477737	1.5940125	92.991381
## 10	1130872	10617423	1.8871093	92.865695
## 11	1239108	1415872	1.3588311	NA
## 12	790760	1787065	1.1055883	59.043119
## 13	2050953	12671821	1.8456716	97.414496
## 14	1621797	6732219	1.8373960	96.642507
## 15	1653310	3155070	1.5171366	90.442459
## 16	499547	2913314	1.6164663	96.804303
## 17	1553642	4467673	1.5549798	11.505349
## 18	1680185	4648794	2.2211124	NA
## 19	2057327	1344212	1.2450113	NA
## 20	1807444	6045680	2.1081135	2.386026
## 21	3477649	6892503	2.5344854	97.275794
## 22	1497291	9986857	2.1011921	86.750630

## 23	1840052	5639632	1.2694007	98.613609
## 24	978257	2976149	2.3006429	96.826251
## 25	1270979	6137428	1.6153298	89.818822
## 26	1363543	1068778	1.4644792	98.160418
## 27	1580319	1934408	1.0059011	NA
## 28	1154399	3080156	1.6950814	93.873937
## 29	1800922	1359711	1.3787982	98.425521
## 30	1627759	8882190	2.5850258	95.945052
## 31	1758157	2096829	2.1102352	94.752811
## 32	3046613	19453561	2.4844643	91.555487
## 33	1325587	10488084	1.3244616	97.855868
## 34	587528	762062	1.3793228	98.508743
## 35	1174479	11689100	1.8298194	97.572250
## 36	1020992	3956971	1.6111383	97.886550
## 37	1297340	4217737	1.3291127	93.405923
## 38	1146745	12801989	2.2843813	96.636940
## 39	137018	3386941	0.9261847	43.996313
## 40	1597981	5148714	1.6453565	96.145957
## 41	554594	884659	1.6358983	98.221292
## 42	1204704	6829174	1.4484229	98.253213
## 43	1110234	28995881	1.7513025	96.441091
## 44	1603589	3205958	0.5715271	97.608302
## 45	2790193	623989	1.0556465	98.473813
## 46	1205933	8535519	1.6778194	8.409783
## 47	1007692	7614893	1.3209238	49.777071
## 48	1691044	1792147	1.7651783	97.592883
## 49	619044	5822434	1.2022435	NA
## 50	1368012	578759	1.1970607	97.900760

Now we summarise our data using summary method, to check the max, min, mean, median etc values of different columns.

```
summary(f_data)
```

## State	Total.Cases	Total.Deaths	Total.Recovered
## Length:50	Min. : 24440	Min. : 258	Min. : 11041
## Class :character	1st Qu.: 206938	1st Qu.: 2428	1st Qu.: 131916
## Mode :character	Median : 461206	Median : 7421	Median : 458815
##	Mean : 675425	Mean : 12067	Mean : 607725
##	3rd Qu.: 840571	3rd Qu.: 13784	3rd Qu.: 822844
##	Max. : 3828483	Max. : 63758	Max. : 2903306
##			
## Active.Cases	Total.Cases.1.mil.population	Death.1.mil.population	
## Min. : 115	Min. : 26924	Min. : 366	
## 1st Qu.: 2380	1st Qu.: 95636	1st Qu.: 1287	
## Median : 13175	Median : 106950	Median : 1756	
## Mean : 100356	Mean : 100993	Mean : 1663	
## 3rd Qu.: 100356	3rd Qu.: 112755	3rd Qu.: 2056	
## Max. : 1687167	Max. : 145367	Max. : 2982	
##			
## Total.Tests	Tests.1.mil.population	Population	Percent_Death
## Min. : 447733	Min. : 137018	Min. : 578759	Min. : 0.5388
## 1st Qu.: 2781684	1st Qu.: 1043302	1st Qu.: 1975013	1st Qu.: 1.3256

```
## Median : 5178674 Median :1344565 Median : 4558234 Median :1.6159
## Mean : 9973373 Mean :1439948 Mean : 6597227 Mean :1.6302
## 3rd Qu.:10924996 3rd Qu.:1688329 3rd Qu.: 7530849 3rd Qu.:1.8767
## Max. :70392011 Max. :3477649 Max. :39512223 Max. :2.5850
##
## Percent.Recovered
## Min. : 2.386
## 1st Qu.:91.277
## Median :96.580
## Mean :85.658
## 3rd Qu.:97.659
## Max. :98.614
## NA's :6
```

We make important observations like mean death % is around 1.6 and over 85 % of all patients have recovered.  
we now check the total cases, total active cases, total recovered cases and total deaths across USA

```
sum(f_data$Total.Cases)
```

```
## [1] 33771259
```

```
sum(f_data$Active.Cases)
```

```
## [1] 5017801
```

```
sum(f_data$Total.Recovered)
```

```
## [1] 30386227
```

```
sum(f_data$Total.Deaths)
```

```
## [1] 603373
```

Similarly, we now check the mean cases, mean active cases, mean recovered cases and mean deaths across USA.

```
mean(f_data$Total.Cases)
```

```
## [1] 675425.2
```

```
mean(f_data$Active.Cases)
```

```
## [1] 100356
```

```
mean(f_data$Total.Recovered)
```

```
## [1] 607724.5
```

```
mean(f_data$Total.Deaths)
```

```
## [1] 12067.46
```

We now check the top 5 and least 5 states in terms of active cases and total deaths.

```
f_data %>% rownames_to_column() %>% top_n(-5, Active.Cases) %>% pull("State","Active.Cases")
```

```
##           428           195           124           178           115
## "Montana" "New Hampshire" "North Dakota" "South Dakota" "Vermont"
```

```
f_data %>% rownames_to_column() %>% top_n(5, Active.Cases) %>% pull("State","Active.Cases")
```

```
##      1687167      405351      441940      612842      223200
## "California" "Kentucky" "Maryland" "Virginia" "Washington"
```

```
f_data %>% rownames_to_column() %>% top_n(-5, Total.Deaths) %>% pull("State","Total.Deaths")
```

```
##       370       518       861       258       751
## "Alaska" "Hawaii" "Maine" "Vermont" "Wyoming"
```

```
f_data %>% rownames_to_column() %>% top_n(5, Total.Deaths) %>% pull("State","Total.Deaths")
```

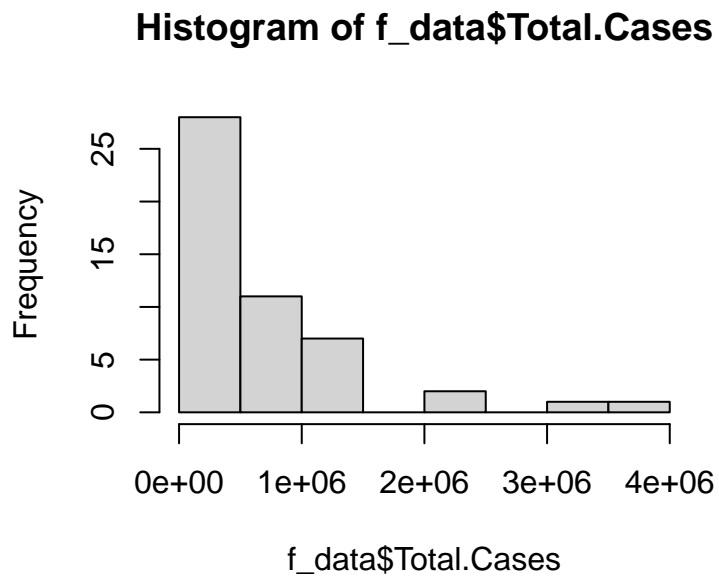
```
##       63758       37985       54065       27821       52722
## "California" "Florida" "New York" "Pennsylvania" "Texas"
```

Here we can observe that California has most number of active cases, and has had maximum deaths. While Montana has least no of active cases, and Alaska he least no of total deaths.

## 2 PLOTS

### 2.1 Histograms

```
hist(f_data$Total.Cases)
```

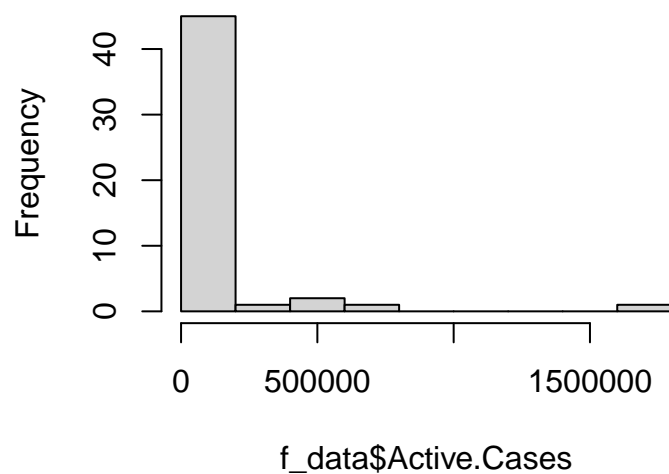


in the above figure we see frequency of states having total cases in distinct half million ranges. NOTE:  $1e+06 = 1000000$

we will now observe some other histograms on available stats.

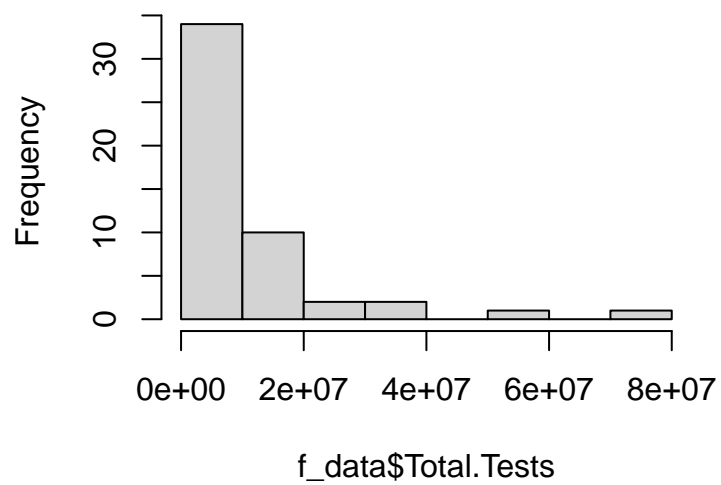
```
hist(f_data$Active.Cases)
```

**Histogram of f\_data\$Active.Cases**



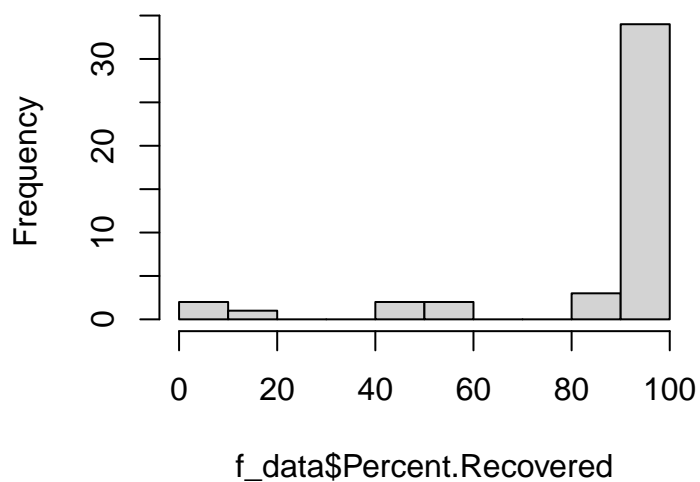
```
hist(f_data$Total.Tests)
```

**Histogram of f\_data\$Total.Tests**



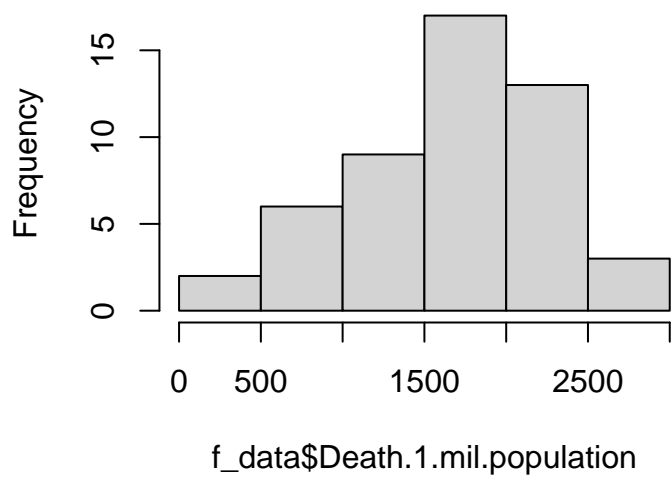
```
hist(f_data$Percent.Recovered)
```

**Histogram of f\_data\$Percent.Recovered**



```
hist(f_data$Death.1.mil.population)
```

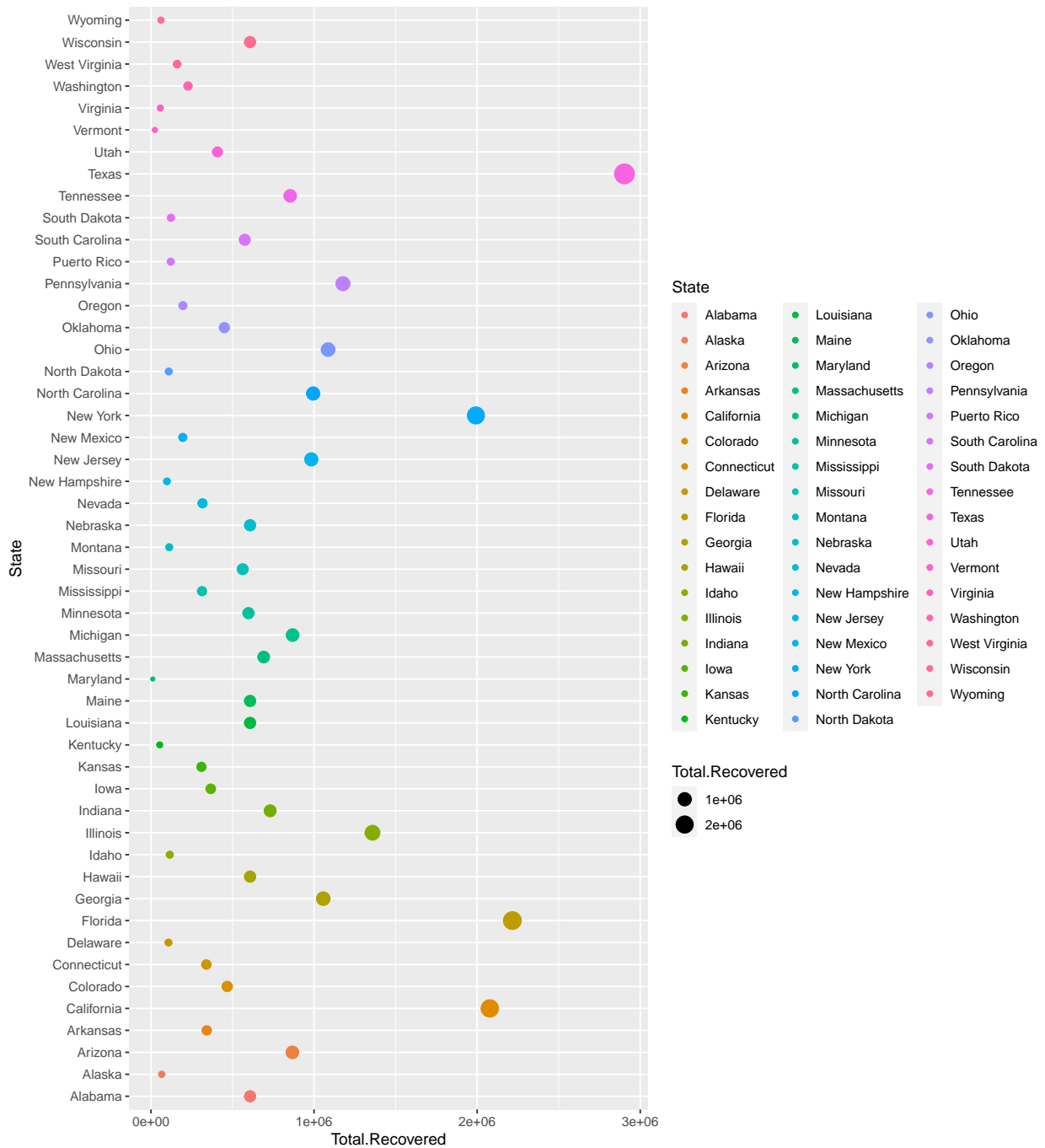
**Histogram of f\_data\$Death.1.mil.population**



## 2.2 GGplots

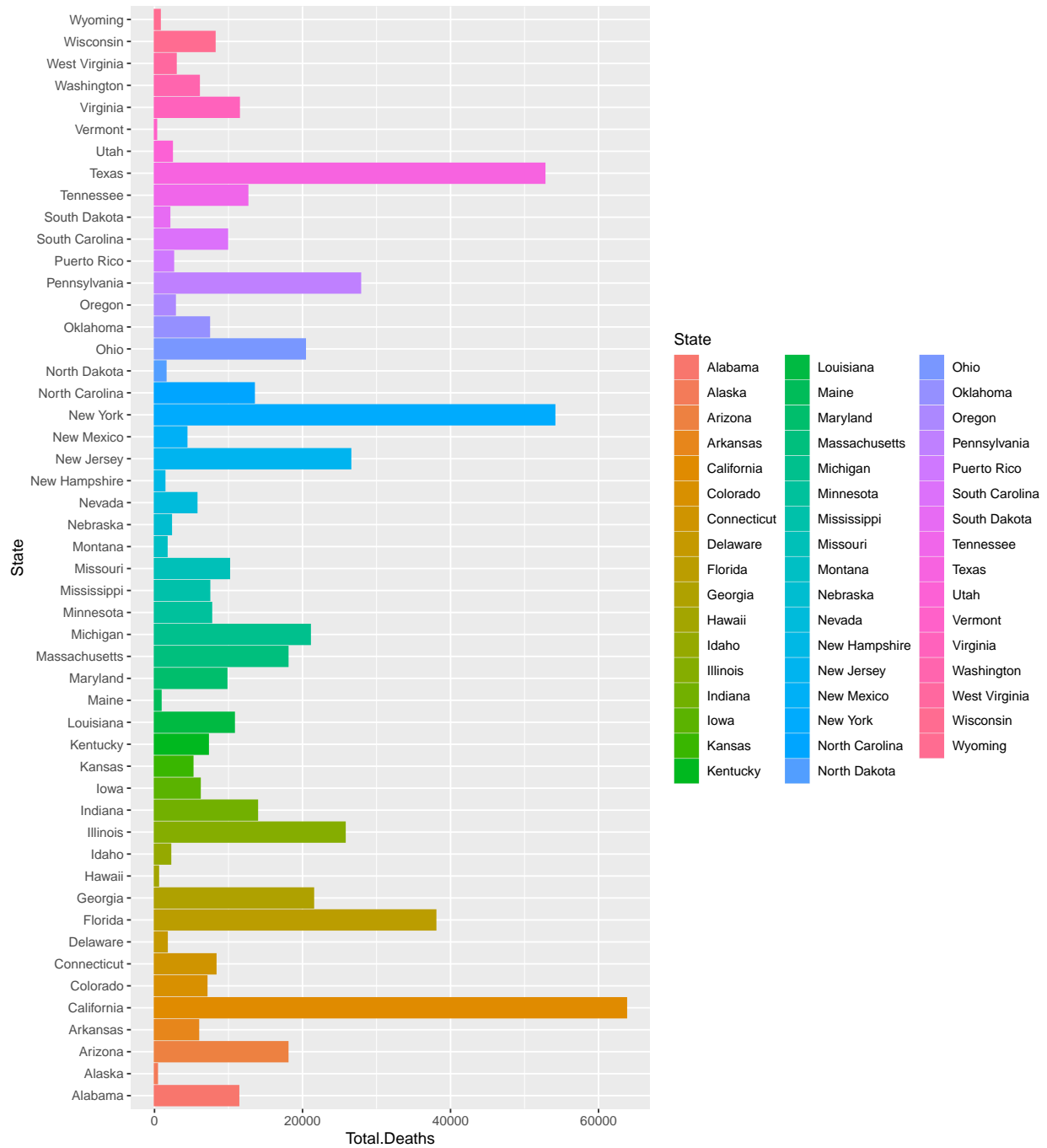
Next we move on to observe several state wise stats using ggplot. we will use bar representation and point representation alternatively to observe.

```
ggplot(f_data, aes(x = Total.Recovered, y = State, col = State, size = Total.Recovered)) + geom_point()
```

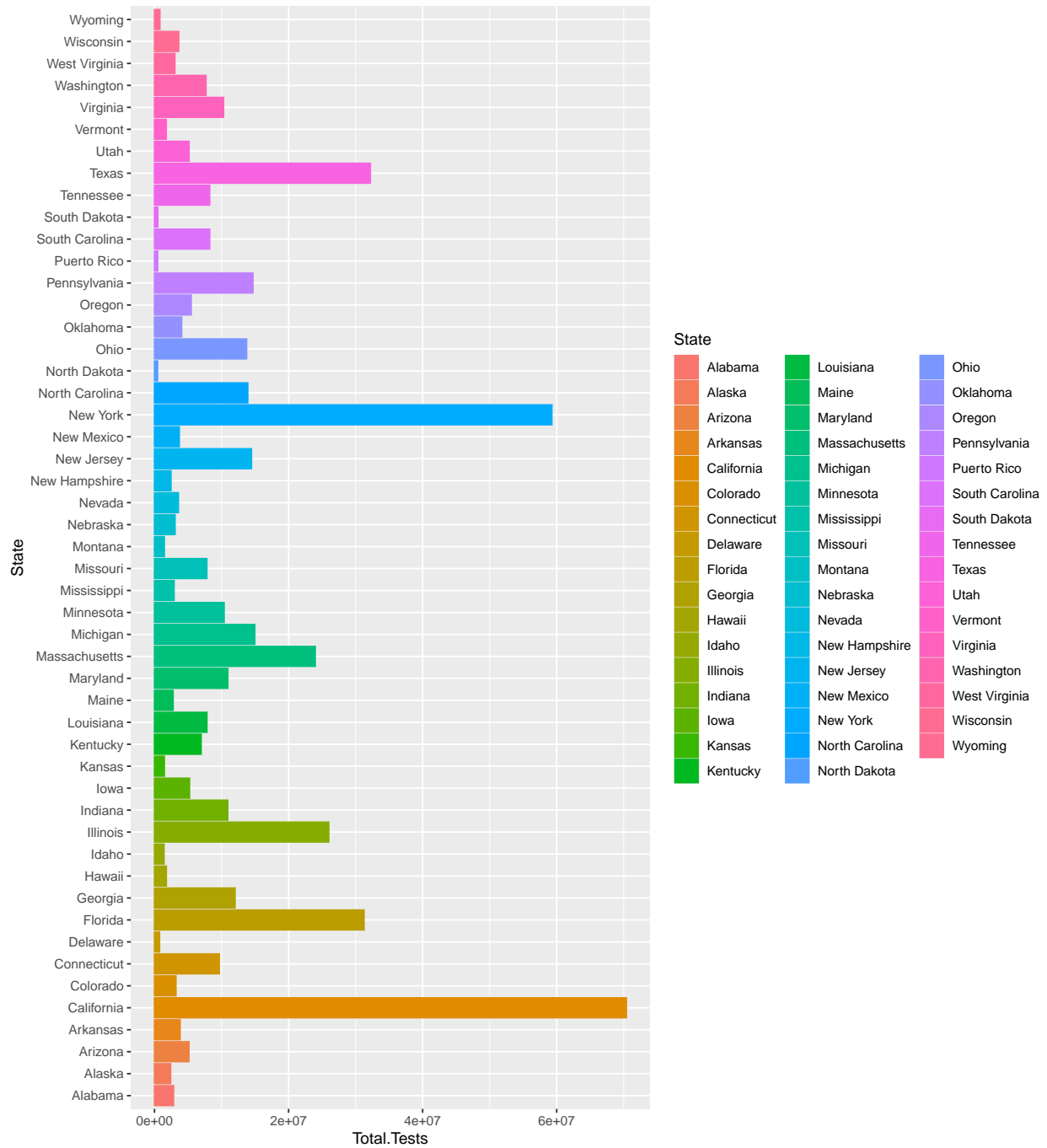


```
ggplot(f_data, aes(x = Total.Deaths, y = State, col = State, fill = State)) + geom_bar(stat = "identity")
```

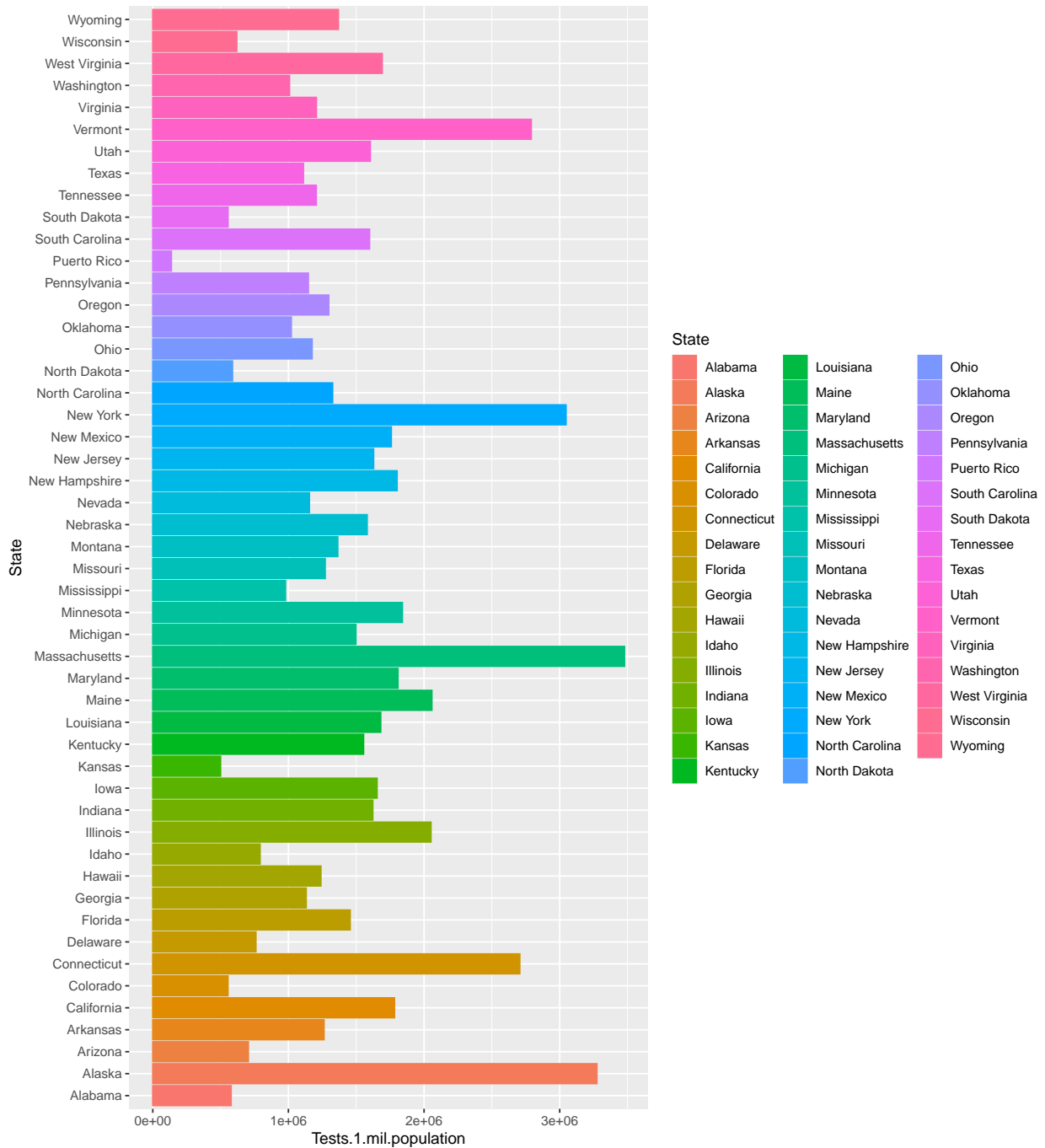




```
ggplot(f_data, aes(x = Total.Tests, y = State, col = State, fill = State)) + geom_bar(stat = "identity")
```



```
ggplot(f_data, aes(x = Tests.1.mil.population, y = State, col = State, fill = State)) + geom_bar(stat =
```



Next we see the percentage deaths and recovered in each state

```
ggplot(f_data, aes(x = Percent_Death, y = State, col = State, size = Percent_Death)) + geom_point()
```

