The City University of New York School of Professional Studies Knowledge and Visual Analytics (DATA 608)

Data Insights to Save the life of U.S. Military Personnel

Alexis Mekueko

12/15/2021

Github Link
Web Link

Benefits

The interest of exploring U.S. Military death data is visualize these death to the public so that something can be done to reduce this death. We all know that U.S. Military involves politicians, technology, industry, healthcare and government. Thus, by displaying this data to the public, all these entities can contribute each at their power level to take major decisions that could end up saving more lives in military. These decisions can be to improving military mechanics, to helping politicians to make better policy, to adjusting military strategy, to doctors and paramedical to rethink and find appropriate health-plan for military personnel. I plan to become a consultant using my skills as data scientist in various domain of the society to present meaningful report to government entities, companies, and organizations to help them in decision making. So, this project will contribute to building skills necessary for one to be successful in data science.

Research questions

What is the death rate of military personnel over the course of 20 years?
What is the death rate of military personnel in active duty of the course of 20 years?
What is the death ration of military personnel by accident and illness?
Do military personnel dies more by homicide than combat?
Do military personnel die more by illness than accident?

Data source

We were looking at open-source data like kaggle.com and found some interesting dataset about military that no one has not made a any contribution on it. The original source of the dataset ('ActiveDutyDeathNo') is from: Defense Casualty Analysis System (DCAS) , https://dcas.dmdc.osd.mil/dcas/pages/report_by_year_manner.xhtml. Data is completely free and represents 20 years (1980-2010) of data collected on U.S. Active Duty Military Deaths. The details of the dataset can be seen below:

Loading Data

-The dataset is pull out from Github in a csv file into Rstudio. We will use R programming language to manipulate and visualize the dataset.
-In addition, will explore the possibility to use python programming language to build a shiny app.

Scrubing Data

Tidying up data

```
## 'data.frame':    31 obs. of  14 variables:
##  $ year                          : int  1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 ...
##  $ active.duty                   : int  2050758 2093032 2112609 2123909 2138339 2150379 2177845 2166611 2121659 2112128 ...
##  $ full.time..est..guard.reserve: int  22000 22000 41000 49000 55000 64000 69000 71000 72000 74200 ...
##  $ selected.reserve.fte          : int  86872 91719 97458 100455 104583 108806 113010 115086 115836 117056 ...
##  $ total.military.fte            : int  2159630 2206751 2251067 2273364 2297922 2323185 2359855 2352697 2309495 2303384 ...
##  $ total.deaths                  : int  2392 2380 2319 2465 1999 2252 1984 1983 1819 1636 ...
##  $ accident                      : num  1556 1524 1493 1413 1293 ...
##  $ hostile.action                : int  0 0 0 18 1 0 2 37 0 23 ...
##  $ homicide                      : int  174 145 108 115 84 111 103 104 90 58 ...
##  $ illness                       : int  419 457 446 419 374 363 384 383 321 294 ...
##  $ pending                       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ self.inflicted                : int  231 241 254 218 225 275 269 260 285 224 ...
##  $ terrorist.attack              : int  1 0 2 263 6 5 0 2 17 0 ...
##  $ undetermined                  : int  11 13 16 19 16 22 27 25 26 37 ...
```

##Organizing data

```
-Checking for missing values
-Checking for empty values
```

```
##
## The dataset contains missing values for a total record of :  0
```

```
##
## The dataset contains empty values for a total record of :  FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

Visualizing Empty and Missing Values

Exploring Data

```
-Inspect data and understand the characteristic of the data Looking for relationship, patterns and values,
```

```
##
## Attaching package: 'gridExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

```
##
## Attaching package: 'kableExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##     group_rows
```

U.S.Active Duty Military Deaths, 1980-2010

year	active.duty	full.time..est..guard.reserve	selected.reserve.fte	total.military.fte	total.deaths	accident	hostile.action	homicide	illness	pending	self.inflicted	terrorist.attack	undetermined
1980	2050758	22000	86872	2159630	2392	1556	0	174	419	0	231	1	11
1981	2093032	22000	91719	2206751	2380	1524	0	145	457	0	241	0	13
1982	2112609	41000	97458	2251067	2319	1493	0	108	446	0	254	2	16
1983	2123909	49000	100455	2273364	2465	1413	18	115	419	0	218	263	19
1984	2138339	55000	104583	2297922	1999	1293	1	84	374	0	225	6	16
1985	2150379	64000	108806	2323185	2252	1476	0	111	363	0	275	5	22
1986	2177845	69000	113010	2359855	1984	1199	2	103	384	0	269	0	27
1987	2166611	71000	115086	2352697	1983	1172	37	104	383	0	260	2	25
1988	2121659	72000	115836	2309495	1819	1080	0	90	321	0	285	17	26
1989	2112128	74200	117056	2303384	1636	1000	23	58	294	0	224	0	37
1990	2046806	74250	137268	2258324	1507	880	0	74	277	0	232	1	43
1991	1943937	70250	184002	2198189	1787	931	147	112	308	0	256	0	33
1992	1773996	67850	111491	1953337	1293	676	0	109	252	0	238	1	17
1993	1675269	68500	105768	1849537	1213	632	0	86	221	0	236	29	9
1994	1581649	65000	99833	1746482	1075	544	0	83	206	0	232	0	10
1995	1502343	65000	94585	1661928	1040	538	0	67	174	0	250	7	4
1996	1456266	65000	92409	1613675	974	527	1	52	173	0	188	19	14
1997	1418773	65000	94609	1578382	817	433	0	42	170	0	159	0	13
1998	1381034	65000	92536	1538570	827	445	0	26	174	0	165	3	14
1999	1367838	65000	93104	1525942	796	439	0	38	154	0	150	0	15
2000	1372352	65000	93078	1530430	832	429	0	37	180	0	153	17	16
2001	1384812	65000	102284	1552096	943	461	12	49	197	0	153	46	25
2002	1411200	66000	149942	1627142	1051	565	17	54	213	0	174	0	28
2003	1423348	66000	243284	1732632	1399	597	312	46	231	1	190	0	22
2004	1411287	66000	234629	1711916	1847	605	735	46	256	0	197	0	8
2005	1378014	66000	220000	1664014	1929	646	739	54	280	1	182	0	27
2006	1371533	72000	168000	1611533	1882	561	769	47	257	8	213	0	27
2007	1368226	72000	168000	1608226	1953	561	847	52	237	22	211	0	23
2008	1402227	73000	207917	1683144	1440	506	352	47	244	6	259	1	25
2009	1421668	75000	144083	1640751	1515	467	346	77	277	19	302	0	27
2010	1430985	76000	178193	1685178	1485	424	456	39	238	22	289	0	17

```
##
##
## |                                |usM (N = 31)                          |
## |:-------------------------------|:-------------------------------------|
## |**year**                        |                            |
## |   minimum            |1,980                                 |
## |   median (IQR)       |1,995 (1,987.50, 2,002.50)            |
## |   mean (sd)          |1,995.00 &plusmn; 9.09                |
## |   maximum            |2,010                                 |
## |**active.duty**                 |                            |
## |   minimum            |1,367,838                             |
## |   median (IQR)       |1,502,343 (1,406,713.50, 2,102,580.00)|
## |   mean (sd)          |1,702,284.90 &plusmn; 337,422.87      |
## |   maximum            |2,177,845                             |
## |**full.time..est..guard.reserve**|                           |
## |   minimum            |22,000                                |
## |   median (IQR)       |66,000 (65,000.00, 71,500.00)         |
## |   mean (sd)          |63,614.52 &plusmn; 13,263.52          |
## |   maximum            |76,000                                |
## |**selected.reserve.fte**        |                            |
## |   minimum            |86,872                                |
## |   median (IQR)       |111,491 (96,033.50, 158,971.00)       |
## |   mean (sd)          |131,157.94 &plusmn; 46,394.45         |
## |   maximum            |243,284                               |
## |**total.military.fte**          |                            |
## |   minimum            |1,525,942                             |
## |   median (IQR)       |1,732,632 (1,620,408.50, 2,254,695.50)|
## |   mean (sd)          |1,897,057.35 &plusmn; 318,690.92      |
## |   maximum            |2,359,855                             |
## |**total.deaths**                |                            |
## |   minimum            |796                                   |
## |   median (IQR)       |1,515 (1,063.00, 1,968.00)            |
## |   mean (sd)          |1,575.29 &plusmn; 526.56              |
## |   maximum            |2,465                                 |
## |**accident**                    |                            |
## |   minimum            |424.00                                |
## |   median (IQR)       |605.00 (516.50, 1,126.00)             |
## |   mean (sd)          |808.81 &plusmn; 391.40                |
## |   maximum            |1,556.00                              |
## |**hostile.action**              |                            |
## |   minimum            |0                                     |
## |   median (IQR)       |1 (0.00, 229.50)                      |
## |   mean (sd)          |155.29 &plusmn; 272.07                |
## |   maximum            |847                                   |
## |**homicide**                    |                            |
## |   minimum            |26                                    |
## |   median (IQR)       |67 (47.00, 103.50)                    |
## |   mean (sd)          |75.13 &plusmn; 35.19                  |
## |   maximum            |174                                   |
## |**illness**                     |                            |
## |   minimum            |154                                   |
## |   median (IQR)       |256 (209.50, 342.00)                  |
## |   mean (sd)          |276.74 &plusmn; 89.15                 |
## |   maximum            |457                                   |
## |**pending**                     |                            |
## |   minimum            |0                                     |
## |   median (IQR)       |0 (0.00, 0.00)                        |
## |   mean (sd)          |2.55 &plusmn; 6.40                    |
## |   maximum            |22                                    |
```

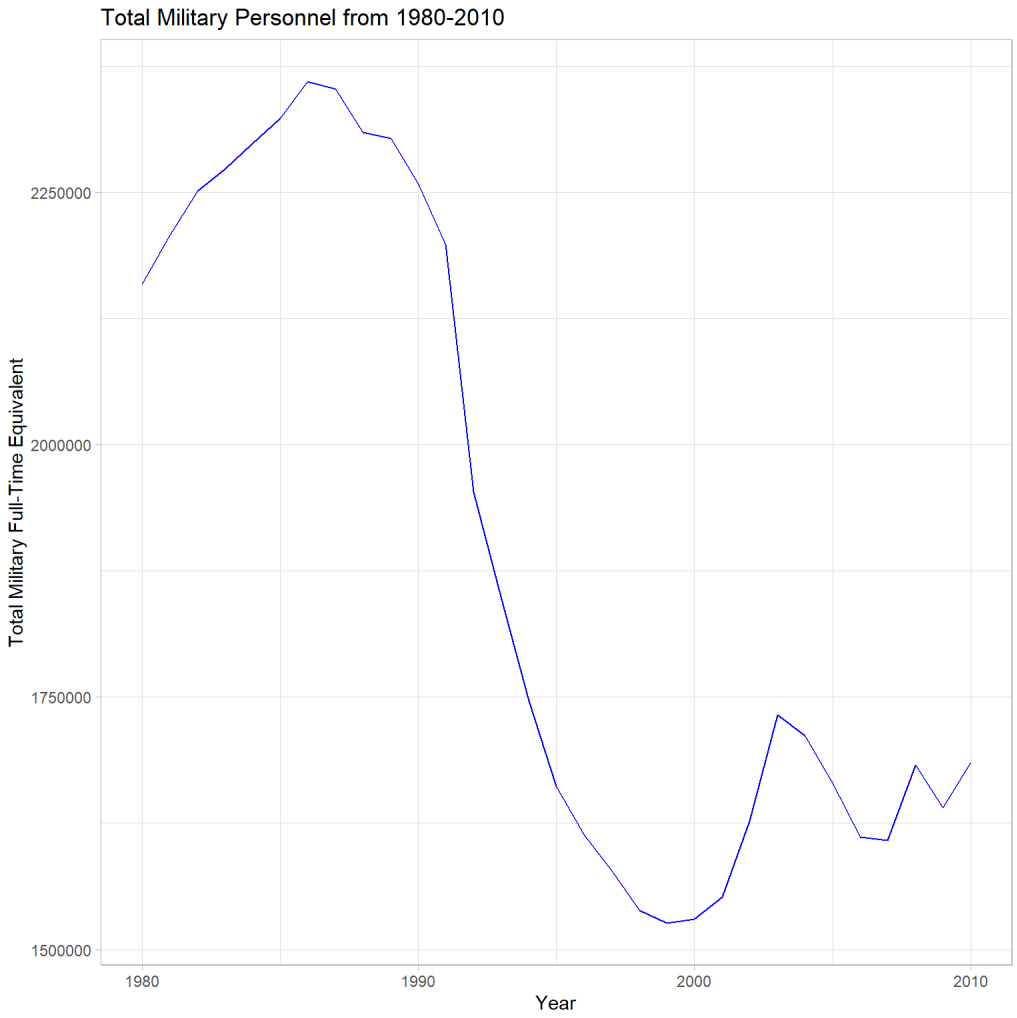
```
## |**self.inflicted**              |                      |
## |   minimum            |150                             |
## |   median (IQR)       |231 (189.00, 255.00)            |
## |   mean (sd)          |222.94 &plusmn; 43.04           |
## |   maximum            |302                             |
## |**terrorist.attack**            |                      |
## |   minimum            |0                               |
## |   median (IQR)       |1 (0.00, 5.50)                  |
## |   mean (sd)          |13.55 &plusmn; 47.44            |
## |   maximum            |263                             |
## |**undetermined**                |                      |
## |   minimum            |4                               |
## |   median (IQR)       |19 (14.00, 26.50)               |
## |   mean (sd)          |20.29 &plusmn; 8.82             |
## |   maximum            |43                              |
```

Looking at the dataset, we can say the dataset is set in 2 parts:

```
-Military Personnel = Active.Duty + Full-Time (est.)Guard-Reserve + Selected.Reserve FTEa + Total.Military.FTE.

-Casualty or Type of Death = Total.Deaths + Accident + Hostile.Action + Homicide + Illness + Pending + Self-Inflicted + Terrorist.Attack + Undetermined

-Definition of death rate: the ratio between deaths and individuals in a specified population during a particular time period :
-The incidence of deaths in a given population during a defined time period (such as one year) that is typically expressed per 1000 or 100,000 individuals.

-Total.Death = Total.Deaths + Accident + Hostile.Action + Homicide + Illness + Pending + Self-Inflicted + Terrorist.Attack + Undetermined

-Total.Military.FTE = Active.Duty + Full-Time (est.)Guard-Reserve + Selected.Reserve FTEa

-Death rate per year = (Total.Death/Total.Military.FTE)*100000

-Death rate% per year = (Total.Death/Total.Military.FTE)*100

-Growth rate in total personnel% per year = (Total.Military.FTE(next_year)-Total.Military.FTE(current_year))*100

-Growth rate in total death % per year = (Total.Death(next_year)-Total.Death(current_year))*100
```

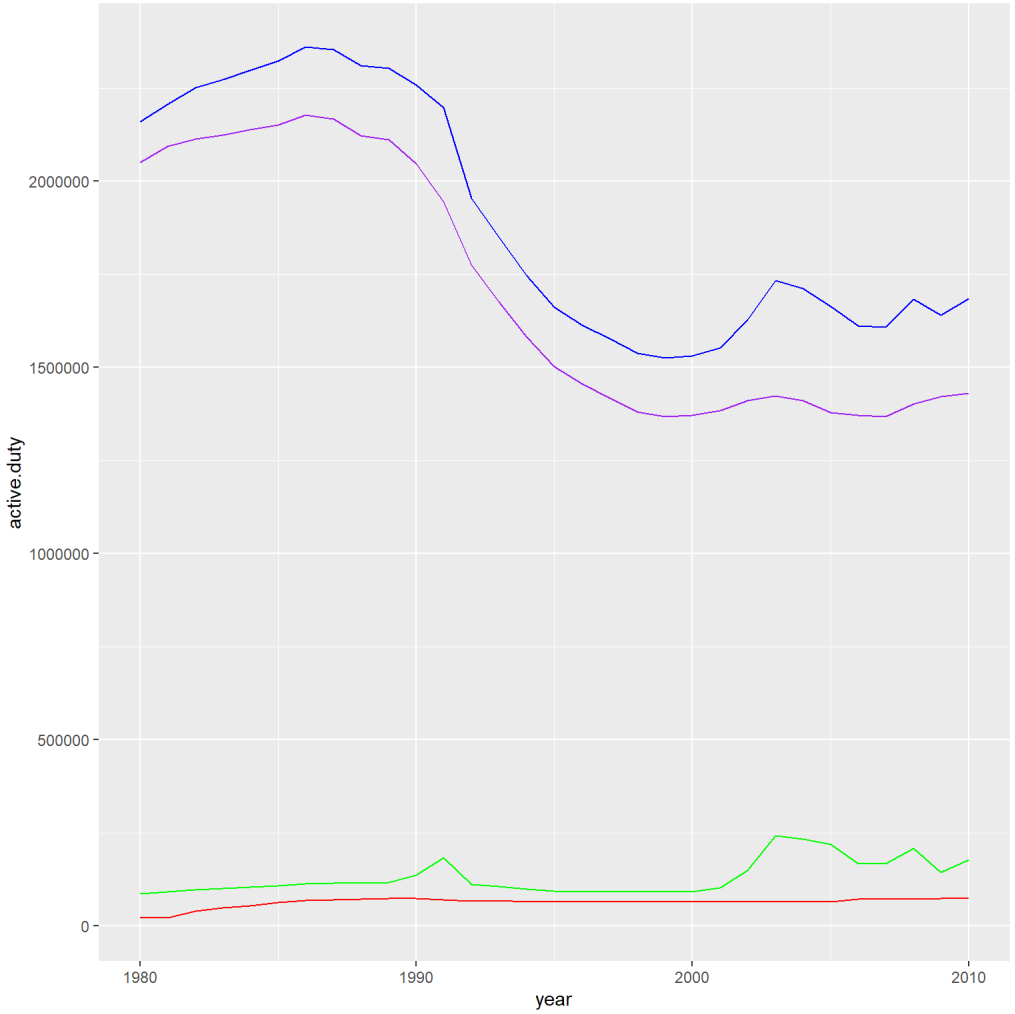
Modelind Data

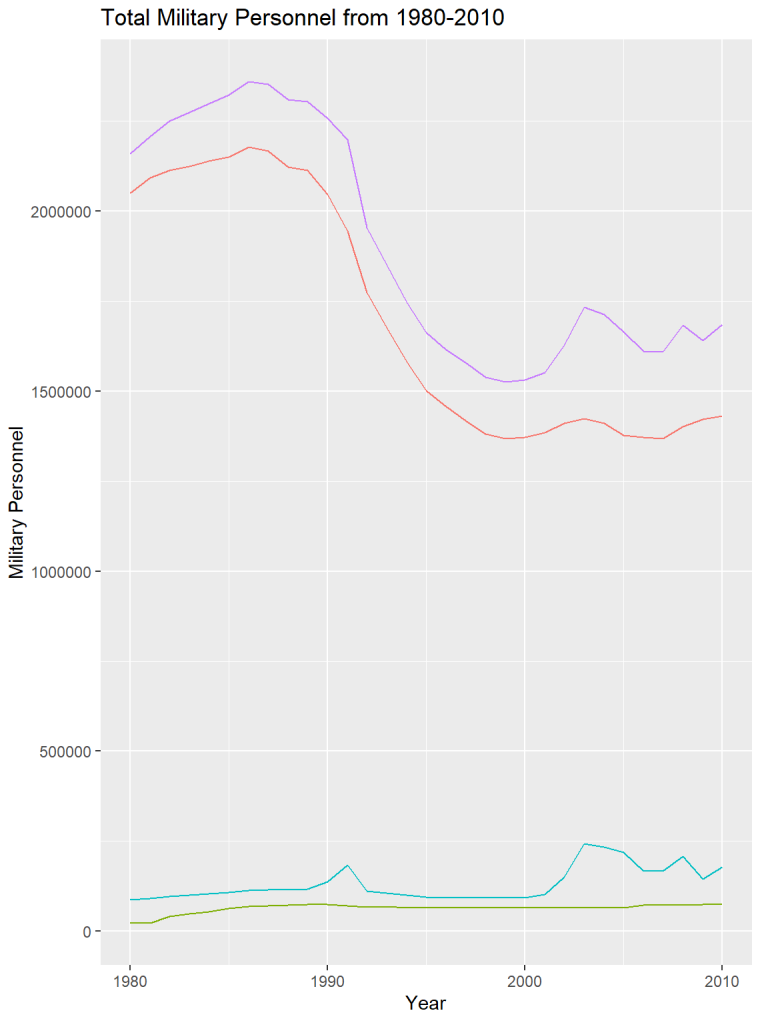
```
-Exploring other data visualization charts
-Explore building apps to display plots: Shiny, Dash
```



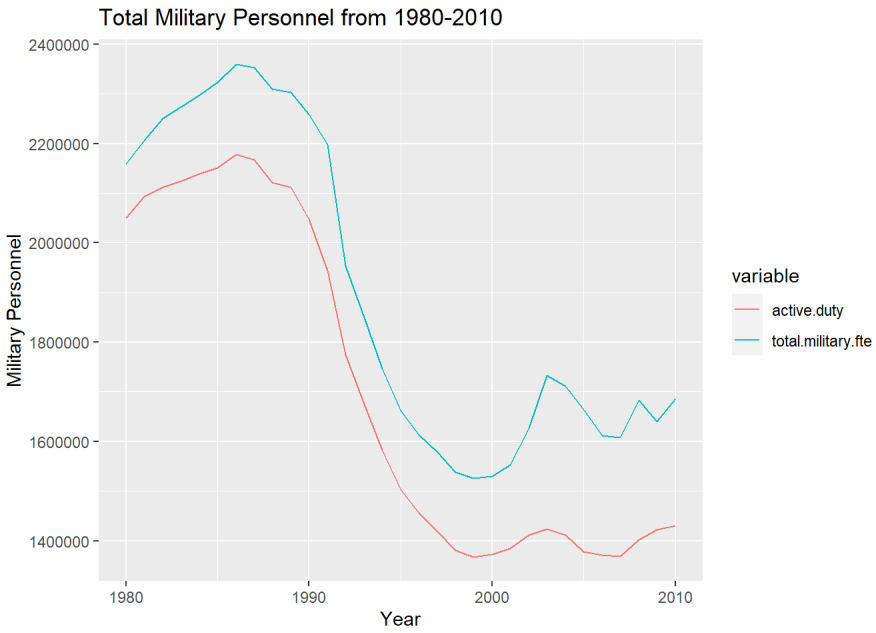

```
##
## Attaching package: 'reshape2'
```

```
## The following object is masked from 'package:tidyr':
##
##     smiths
```


2000000
1500000
1000000
500000
0
active.duty
1980
1990
2000
2010
year



Another way to plot the military personnel against time is to group time block (1980 to 1990 = decade1, 1990 to 2000= decade2, 2000 to 2010 = decade3), while summing the values of other variables within decade. Then, use barplot() or bubble plot. We notice that there is discrepancy active.duty, total.military.fte and full.time..est..guard.reserve,selected.reserve.fte . We can fix this by plotting the 02 variables seperately.

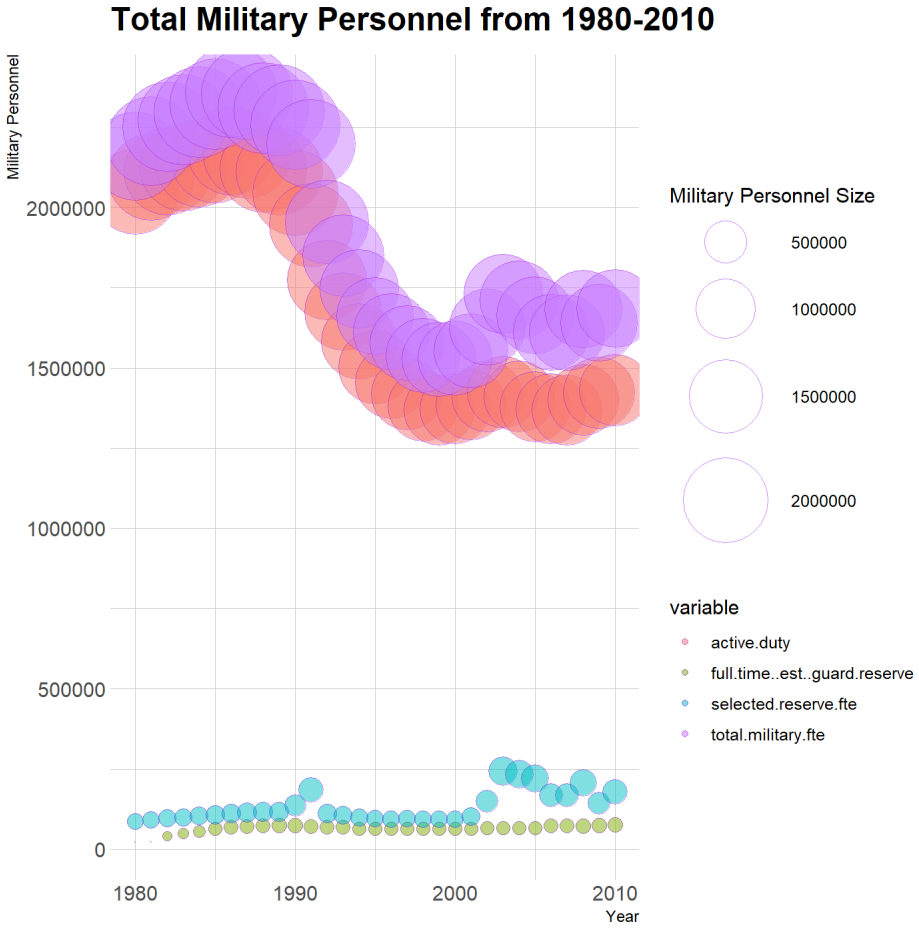



```
## NOTE: Either Arial Narrow or Roboto Condensed fonts are required to use these themes.
```

```
##       Please use hrbrthemes::import_roboto_condensed() to install Roboto Condensed and
```

```
##       if Arial Narrow is not on your system, please see https://bit.ly/arialnarrow
```

```
## Loading required package: viridisLite
```

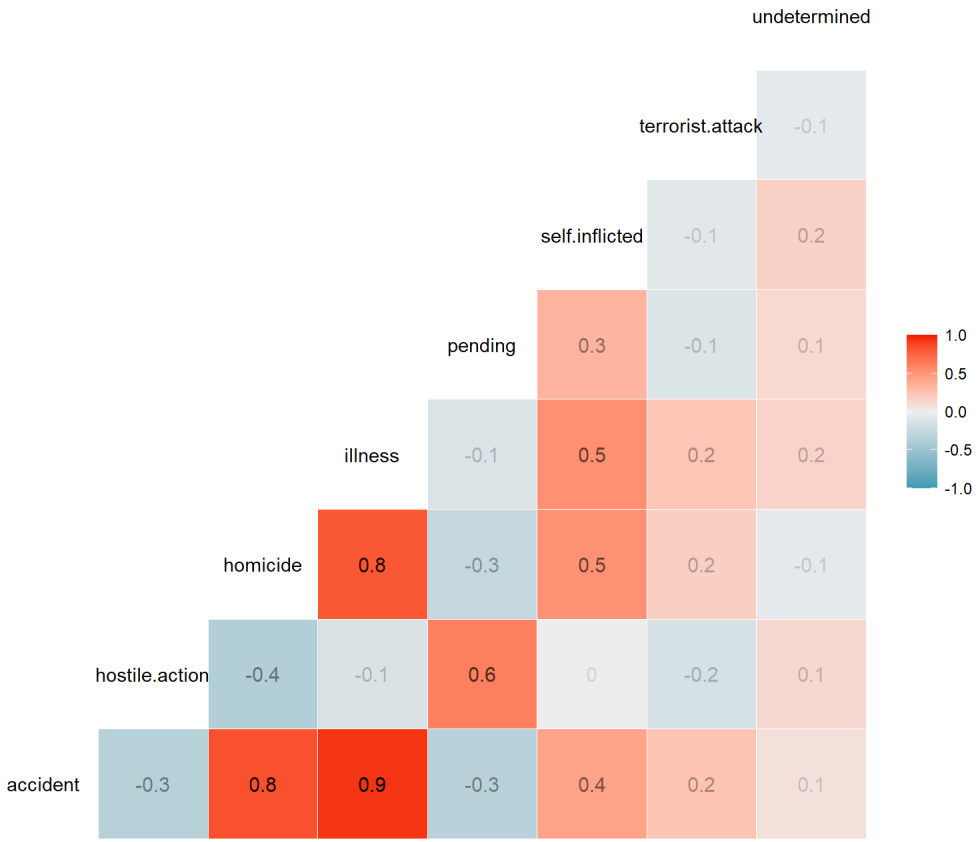


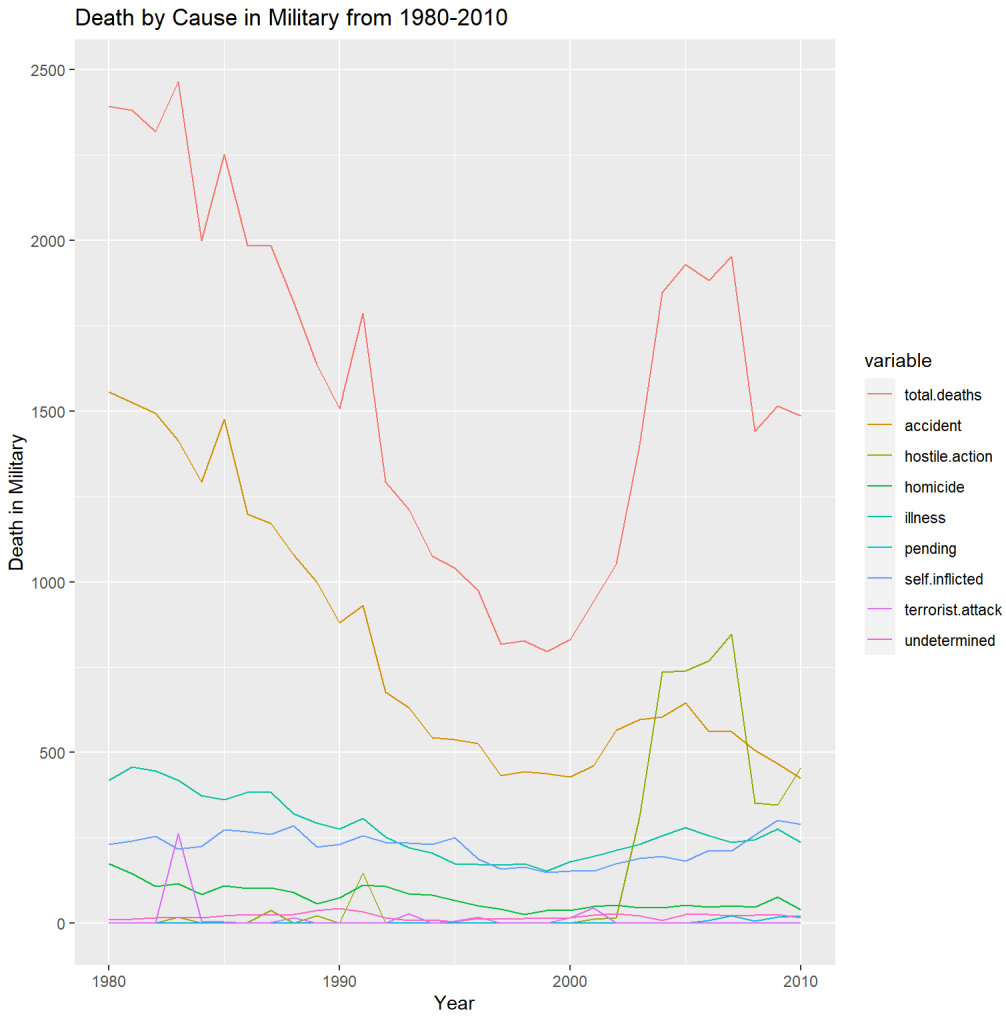
Let's see the military death over the 20 years(1980-2010)

```
##
## Attaching package: 'corrgram'
```

```
## The following object is masked from 'package:plyr':
##
##     baseball
```

```
## The following object is masked from 'package:lattice':
##
##     panel.fill
```


undetermined
terrorist.attack -0.1
self.inflicted -0.1 0.2
pending 0.3 -0.1 0.1
illness -0.1 0.5 0.2 0.2
homicide 0.8 -0.3 0.5 0.2 -0.1
hostile.action -0.4 -0.1 0.6 0 -0.2 0.1
accident -0.3 0.8 0.9 -0.3 0.4 0.2 0.1
1.0
0.5
0.0
-0.5
-1.0




```
##
## Attaching package: 'plotly'
```

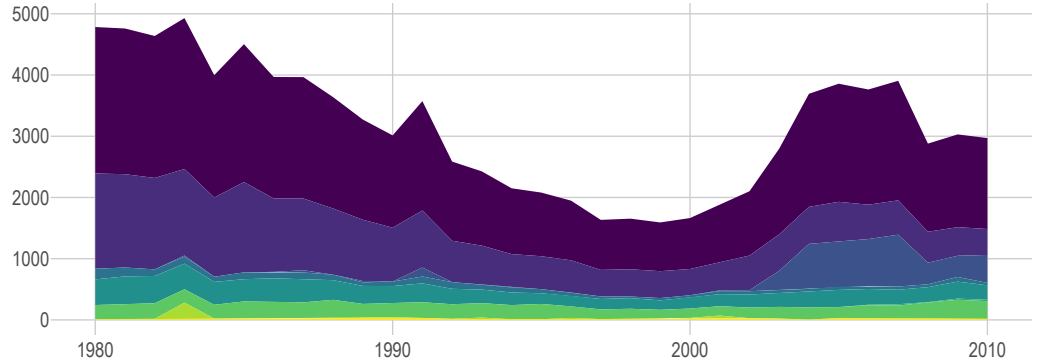
```
## The following objects are masked from 'package:plyr':
##
##     arrange, mutate, rename, summarise
```

```
## The following object is masked from 'package:lessR':
##
##     style
```

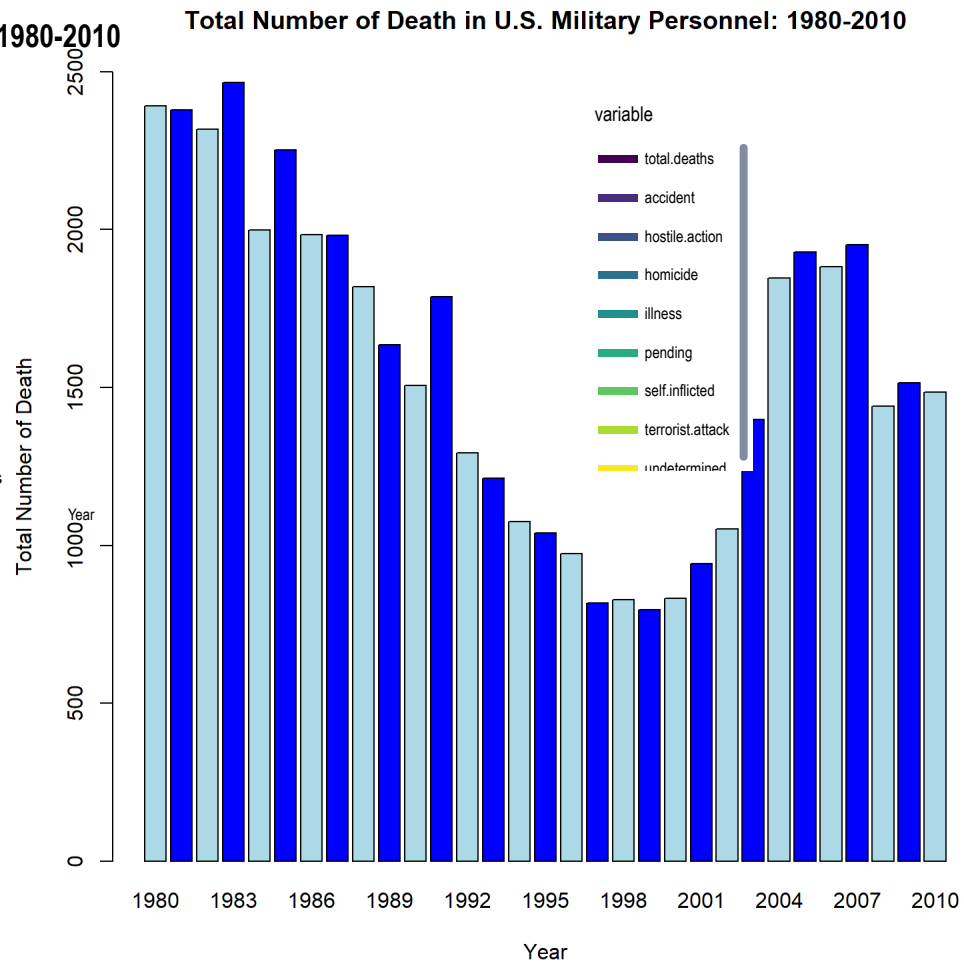
```
## The following object is masked from 'package:ggplot2':
##
##     last_plot
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

```
## The following object is masked from 'package:graphics':
##
##     layout
```



Total.Death = Total.Deaths + Accident + Hostile.Action + Homicide + Illness + Pending + Self-Inflicted + Terrorist.Attack + Undetermined Total.Military.FTE = Active.Duty + Full-Time (est.)Guard-Reserve + Selected.Reserve FTEa



Let's see the total death by year over the course of the 20 years

Some Reference in U.S. Military Battlefield for the past decades:

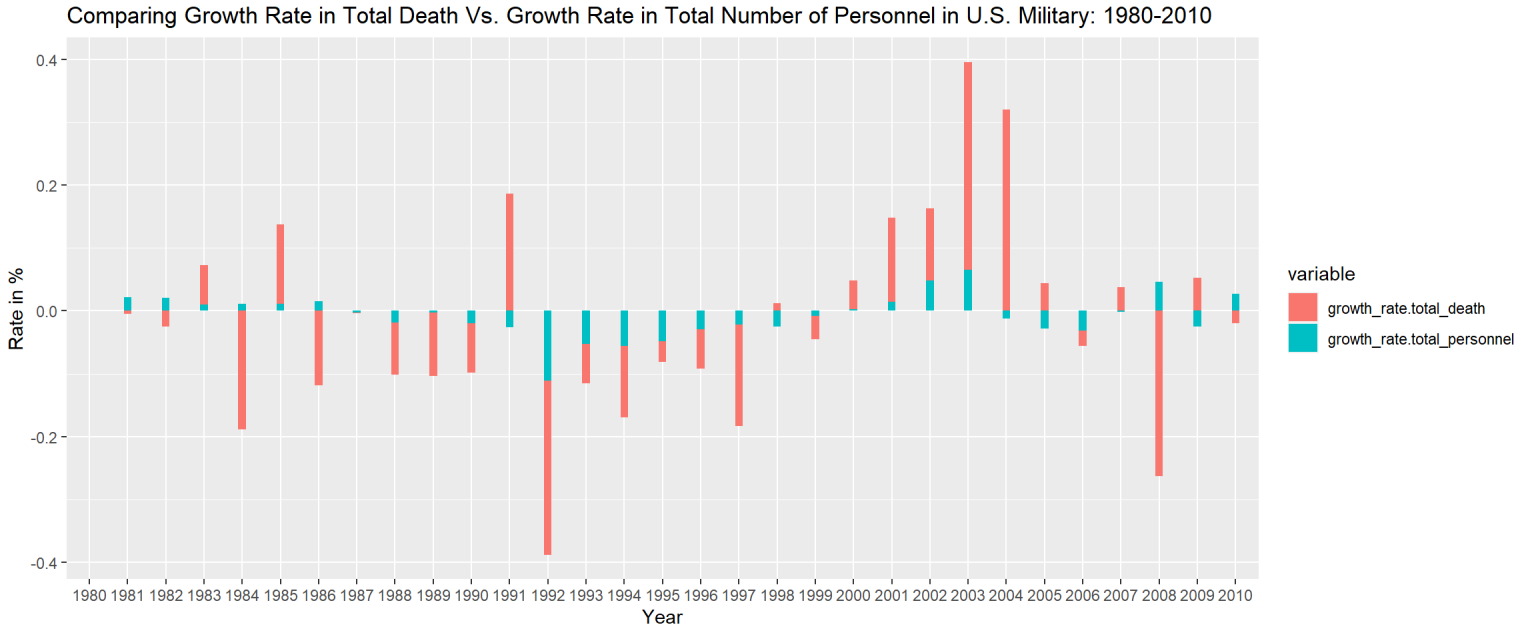
```
1983 - U.S. Military in Grenada
1989 - U.S. Military in Panama
1990 - U.S. Military in Gulf War
1993 - U.S. Military in Somalia War
2001-  U.S in Afghanistan (2001-2021)
2003 - U.S. Military in U.S. Iraq(2003-2011)
```

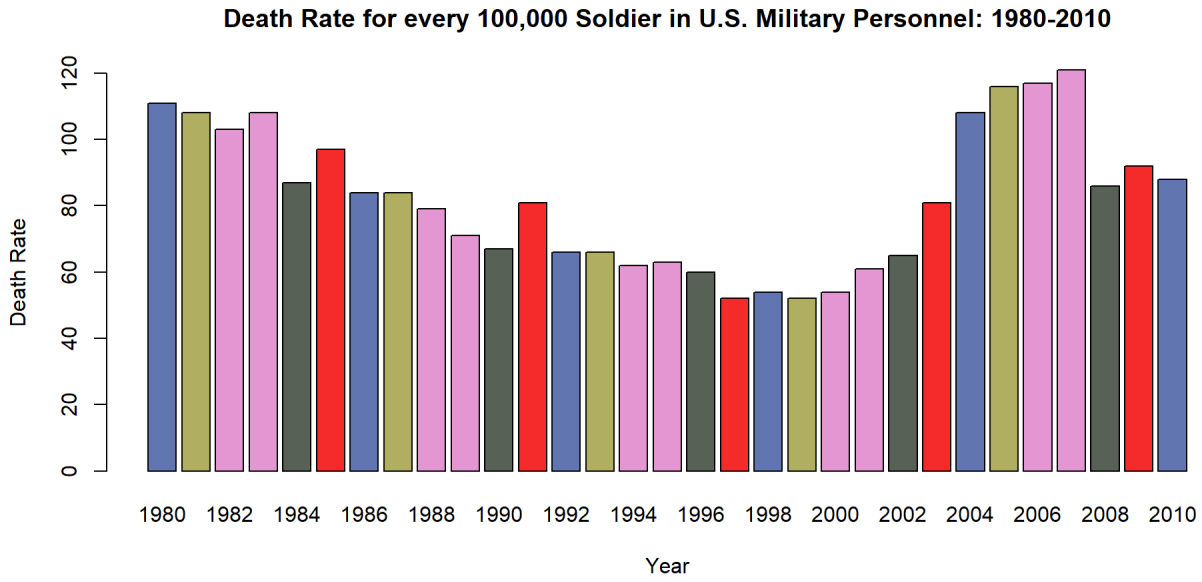
y…	death_rate.accident	death_rate.hostile.action	death_rate.homicide	death_rate.illness
<int>	<dbl>	<dbl>	<dbl>	<dbl>
1980	65.05	0.00	7.27	17.52
1981	64.03	0.00	6.09	19.20

y…	death_rate.accident	death_rate.hostile.action	death_rate.homicide	death_rate.illness
<int>	<dbl>	<dbl>	<dbl>	<dbl>
1982	64.38	0.00	4.66	19.23
1983	57.32	0.73	4.67	17.00
1984	64.68	0.05	4.20	18.71
1985	65.54	0.00	4.93	16.12
1986	60.43	0.10	5.19	19.35
1987	59.10	1.87	5.24	19.31
1988	59.37	0.00	4.95	17.65
1989	61.12	1.41	3.55	17.97

1-10 of 31 rows | 1-5 of 9 columns

Previous 1 2 3 4 Next



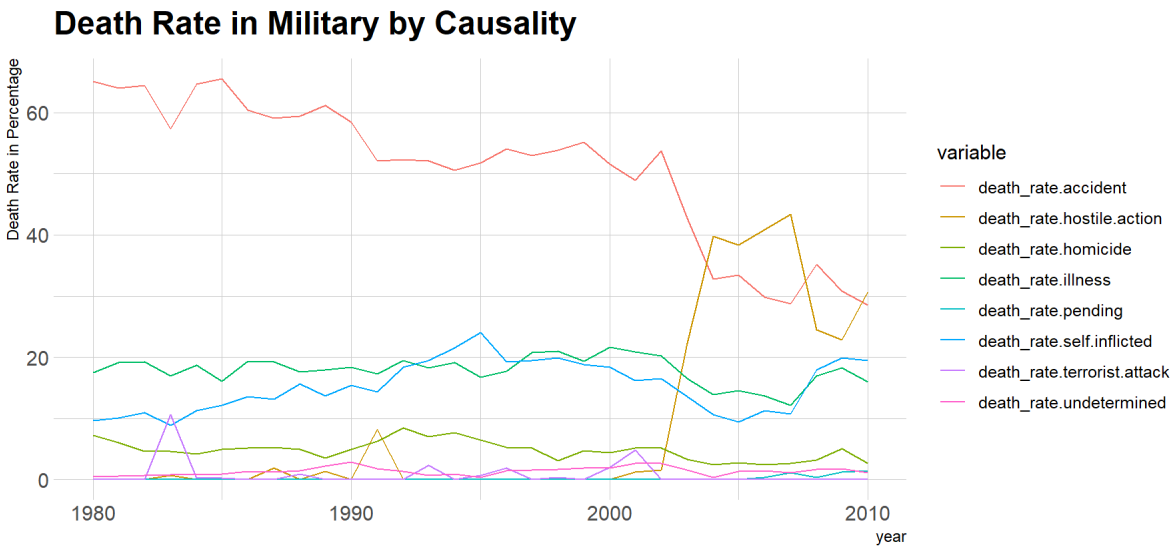


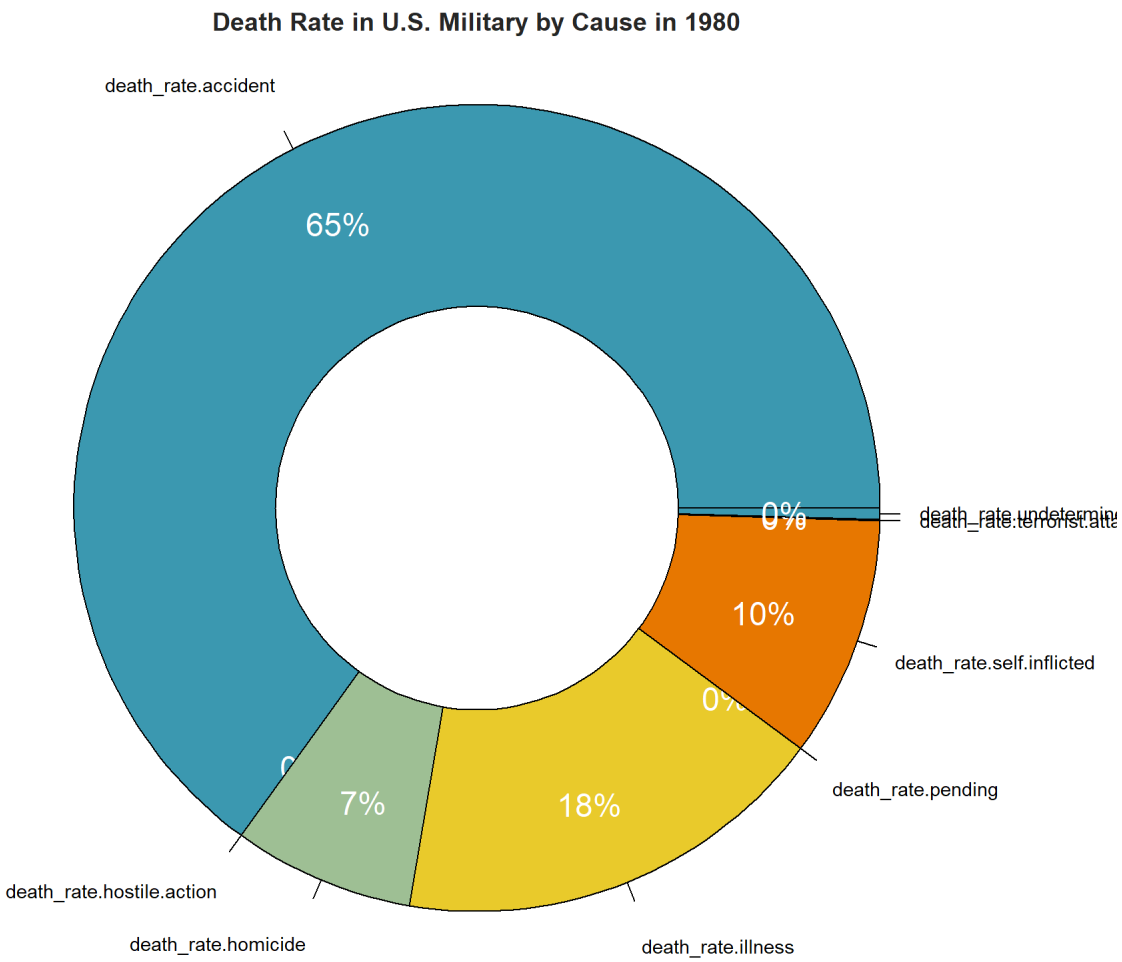
Let's Visualize the differente rates among military personnel.

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v tibble  3.0.6     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
## v purrr   0.3.4
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x plotly::arrange()        masks plyr::arrange(), dplyr::arrange()
## x gridExtra::combine()     masks dplyr::combine()
## x purrr::compact()         masks plyr::compact()
## x plyr::count()            masks dplyr::count()
## x Matrix::expand()         masks tidyr::expand()
## x plyr::failwith()         masks dplyr::failwith()
## x plotly::filter()         masks dplyr::filter(), stats::filter()
## x kableExtra::group_rows() masks dplyr::group_rows()
## x plyr::id()               masks dplyr::id()
## x dplyr::lag()             masks stats::lag()
## x purrr::lift()            masks caret::lift()
## x plotly::mutate()         masks plyr::mutate(), dplyr::mutate()
## x Matrix::pack()           masks tidyr::pack()
## x arules::recode()         masks lessR::recode(), dplyr::recode()
## x plotly::rename()         masks plyr::rename(), dplyr::rename()
## x plotly::summarise()      masks plyr::summarise(), dplyr::summarise()
## x plyr::summarize()        masks dplyr::summarize()
## x Matrix::unpack()         masks tidyr::unpack()
```


Death Rate in Military by Causality
Death Rate in Percentage
60
40
20
0
1980
1990
2000
2010
year
variable
death_rate.accident
death_rate.hostile.action
death_rate.homicide
death_rate.illness
death_rate.pending
death_rate.self.inflicted
death_rate.terrorist.attack
death_rate.undetermined




```
## >>> Suggestions
## PieChart(death.cause, hole=0)  # traditional pie chart
## PieChart(death.cause, values="%")  # display %'s on the chart
## BarChart(death.cause)  # bar chart
## Plot(death.cause)  # bubble plot
## Plot(death.cause, values="count")  # lollipop plot
##
## --- death.rate ---
##
##     n   miss       mean         sd        min        mdn        max
##     8      0     12.500     22.152      0.000      3.865     65.050
```

Interpret Results

```
- There are 03 set of high correlation among the cause of death in military personnel from 1980 to 2010.
    - Illness and Homicide
    - Illness and Accident
    - Homicide and Accident
    - These correlations show that military personnel death by accident and homicide increase with those dying based on illness. In order, the more military     personnel are sick, the likely-hood of more death occurring by accident and homicide.

- These correlations shows also that there is likely more death to occur by homicide when more military personnel die by accident.

- U.S. Military tends to have more casualty when engaging in war.

- Over the course of 20 years(1980-2010) of active duty, U.S. Military has significantly dropped. This might have some explanation with United Nations policy on regulating the size of military of coun
tries around the world.
```

Challenges

-There are few challenges in this project to be overcome:

-Due to the sensibility of the dataset, it can be though to be neutral.

-Rendering the data to a suitable chart was not easy.

-There were a confusion perhaps not in the sense of grammar but more of statistical appreciation of what growth rate is. We thought some information could be reveal while exploring the growth rate among total death in U.S. Military personnel. We wanted to see if this rate was increasing or decreasing from year to year. In addition, let's say a virus is spraying in a population, the rate at which the population is getting contaminated starts at 0 (meaning there were no precedent of such a virus). Therefore, it makes sense to have this rate to eventually settle around zero when the virus is under control and the population is immunized. If we consider this assumption, therefore, it makes no sense to see throughout the timeline of U.S. battlefield the growth rate in total death of military personnel to goes negative. We suspect the formula may need a closer look.

References

https://www.codegrepper.com/code-examples/whatever/insert+image+in+r+markdown

https://sgp.fas.org/crs/natsec/RL32492.pdf

https://cran.r-project.org/web/packages/kableExtra/vignettes/awesome_table_in_html.html

https://www.r-graph-gallery.com/web-line-chart-with-labels-at-end-of-line.html

https://www.r-graph-gallery.com/37-barplot-with-number-of-observation.html

https://sgp.fas.org/crs/natsec/RL32492.pdf

https://bookdown.org/chua/ber642_advanced_regression/r-basics.html

https://forcoast.com/caf/?ses=Y3JlPTE2Mzk2MjMyNTAmdGNpZD1mb3Jjb2FzdC5jb202MWJhYWE1MmNmYjQxNy4zMjA5MTI3NiZ0YXNrPXNlYXJjaCZkb21haW49Zm9yY29hc3QuY29tJmFfaWQ9MyZzZXNzaW9uPWFOZE00YTR6aThfNXppd3RjNUlW&query=Ozn8U1C-qawwWwFYzxvcoNyU_ZefIvSkVrz-gzJO7v-lyko-oh8xCtXU4FIbA