ASSIGNMENT

TECHNOLOGY PARK MALAYSIA

CT127-3-2-PFDA

PROGRAMMING FOR DATA ANALYSIS

APD2F2206IT(FT)

HAND OUT DATE: 25 JULY 2022

HAND IN DATE: 15 AUGUST 2022

WEIGHTAGE: 50%

INSTRUCTIONS TO CANDIDATES:

1 Submit your assignment at the administrative counter.

2 Students are advised to underpin their answers with the use of references (cited using the American Psychological Association (APA) Referencing).

3 Late submission will be awarded zero (0) unless Extenuating Circumstances (EC) are upheld.

4 Cases of plagiarism will be penalized.

5 The assignment should be bound in an appropriate style (comb bound or stapled).

6 Where the assignment should be submitted in both hardcopy and softcopy, the softcopy of the written assignment and source code (where appropriate) should be on a CD in an envelope / CD cover and attached to the hardcopy.

7 You must obtain 50% overall to pass this module.

Table of Contents

Introduction

The data that was given is the employee dataset of a specific company. For this assignment, I am tasked to explore the application of data analytics techniques to the dataset provided. Whilst analysing the data problems related to the datasets and at the same time providing considerable unique properties of the problem domain while testing one or more techniques on it. The original data have 18 columns and a total of 49653 number of rows. In this assignment, I will go through with finding the real issues and problems behind the attrition of employees. The data included each employees data such as birth date and place of work and termination details such as termination reason and status of employee.

Assumption

The dataset given is from the country Canada. The data collected are in sequence of yearly record of employees till current or till terminated.

Import Data

```
#Import data
data=read.csv("C:/Users/User/Desktop/PFDA/employee_attrition.csv",header=TRUE)
```

The figure above shows how dataset are imported into RStudio. The source code include the location of the file and assigning of the dataset's header

Import Library

```
#Import library
library(tidyverse)
library(stringi)
library(plotrix)
library(fmsb)
library(RColorBrewer)
library(skimr)
```

The figure above shows how external library are imported into RStudio.

Library	**Functions**
tidyverse	A collection of R packages built for data research, all of which share a same design philosophy, language, and data format.
stringi	A set of string, text, and natural language processing tools for searching patterns.
plotrix	Tools for data visualisation in R.
fmsb	provide resources for generating radar charts, commonly known as spider plots. It is used to display the values or scores ascribed to a person across numerous quantitative variables, where each variable correlates to a specific axis.
RColorBrewer	resources that may be used to generate graphs using 8 to 12 pre-made colour palettes.

skimr	To summarise a larger set of statistics than summary() by default, including missing, complete, n, and standard deviation. It also permits the functions to report each data type independently and enables the date, logical, and other forms of handles.

Data Transformation

The process of transforming data entails adjusting its format, structure, or values. In projects requiring data analytics, data may be updated at two data pipeline phases. Data transformation can be constructive (adding, duplicating, and replicating data), destructive (removing fields and records), aesthetic (standardising greetings or street names), or structural (renaming, moving, and combining columns in a database).

```
#Data Transformation
##Renaming the header of the dataset
names(data)=c("Employee_ID","Employee_Record_Date","Birth_Date","Hired_Date","Termination_Date","Age",
              "Year_of_Service","City","Department","Job","Store_Code","Gender_s","Gender_Full",
              "Reason_of_Termination","Type_of_Termination","Year_of_Status","Status","Business_Unit")
```

The figure above shows the source code that allows the alteration of the header of the dataset.

Data Cleaning

In order to improve the quality of your data, data cleansing involves finding and correcting potential data inconsistencies and faults. An error is any measurement value (e.g., reported weight) that does not correlate to the actual value of the variable being measured (e.g., actual weight).

```
#Data Cleaning
##Removing spelling errors and removing empty values
data2=data %>%mutate(Termination_Date= case_when(Termination_Date == "1/1/1900" ~"None",Termination_Date !="1/1/1900" ~Termination_Date),
                     City = case_when(City == "New Westminister" ~ "New Westminster", City != "New Westminister" ~ City),
                     Birth_Month = stri_extract_first(Birth_Date,regex = "\\d+"),
                     Termination_Month = case_when(Termination_Date =="None"~"None",Termination_Date !="None" ~stri_extract_first(Termination_Date,regex = "\\d+")),
                     Hire_Month =stri_extract_first(Hired_Date,regex = "\\d+"),
                     Hire_Year =stri_extract_last(Hired_Date,regex = "\\d+"),
                     Year_of_Status = factor(Year_of_Status))%>%arrange(Employee_ID,Year_of_Service) %>%select(-Gender_s)
```

The figure above shows the source code that allows the alteration to group and alter the data in the dataset. Such data need to be change as there are spelling errors such as the data “New Westminster” is misspelled as “New Westminister”. Other data such as “1/1/1900” is also changed to “None” as “1/1/1900” is the default reply when the data is null.

Data Exploration

Data exploration can also refer to the ad hoc querying or visualisation of data to discover potential relationships or insights that may be hidden within the data, without the requirement to generate prior hypotheses.

```
#Data Exploration
names(data2)

nrow(data2)
ncol(data2)

sum(is.na(data2))

unique(data2$Employee_ID)
unique(data2$Age)
unique(data2$Year_of_Service)
unique(data2$City)
unique(data2$Department)
unique(data2$Job)
unique(data2$Store_Code)
unique(data2$Type_of_Termination)
unique(data2$Year_of_Status)
unique(data2$Business_Unit)

distinct(data2,data2$Reason_of_Termination)

table(data2$Age)
table(data2$Gender_Full)
table(data2$City)
table(data2$Department)

summary(data2)

structure(data2)

table(data2$Termination_Date)#42450
```

Source Code	Output

names(data2)	`> names(data2)`
	`[1] "Employee_ID" "Employee_Record_Date" "Birth_Date"`
	`[4] "Hired_Date" "Termination_Date" "Age"`
	`[7] "Year_of_Service" "City" "Department"`
	`[10] "Job" "Store_Code" "Gender_Full"`
	`[13] "Reason_of_Termination" "Type_of_Termination" "Year_of_Status"`
	`[16] "Status" "Business_Unit" "Birth_Month"`
	`[19] "Termination_Month" "Hire_Month" "Hire_Year"`

These commands help see the title of each header assign to the datasets.

Source Code	Output
nrow(data2)	`> nrow(data2)` `[1] 49653`
ncol(data2)	`> ncol(data2)` `[1] 21`

These commands help see the number of rows and the number of columns that the dataset have.

Source Code	Output
sum(is.na(data2))	`> sum(is.na(data2))` `[1] 0`

This command shows the sum of all not available data which is presented in the dataset.

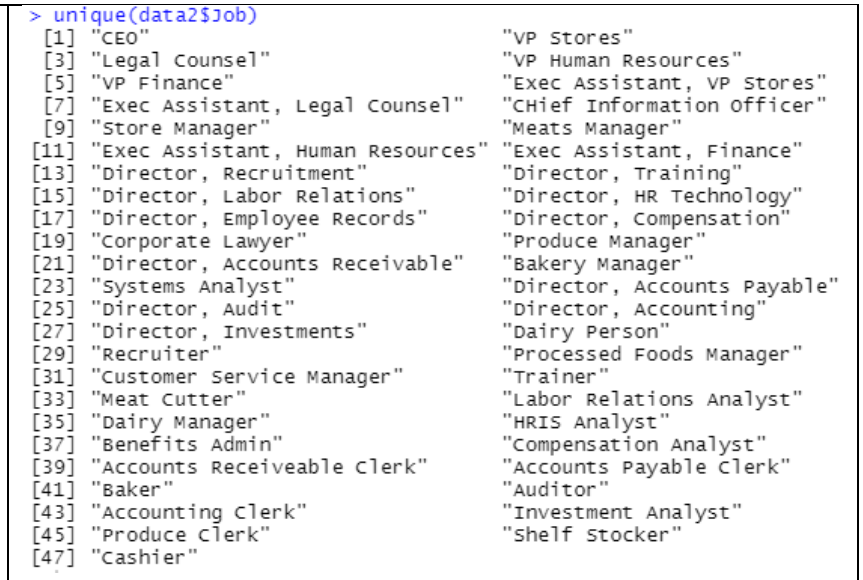
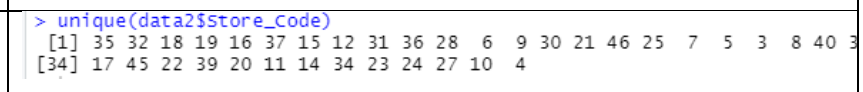
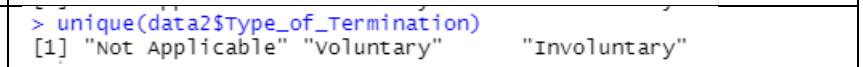
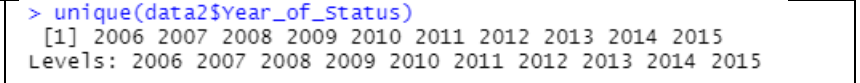
Source Code	Output

unique(data2$Employee_ID)	`> unique(data2$Employee_ID)`
	`[1] 1318 1319 1320 1321 1322 1323 1325 1328 1329 1330 1331 1332 1334 1335 1338 1339 1340`
	`[18] 1341 1343 1344 1346 1347 1351 1352 1353 1355 1357 1358 1359 1360 1362 1363 1365 1366`
	`[35] 1367 1370 1372 1373 1374 1376 1377 1380 1381 1382 1383 1385 1386 1388 1389 1390 1391`
	`[52] 1392 1394 1395 1396 1397 1399 1400 1401 1402 1403 1404 1405 1406 1408 1409 1411 1412`
	`[69] 1413 1414 1417 1419 1421 1422 1425 1427 1428 1430 1431 1433 1434 1435 1436 1438 1439`
	`[86] 1440 1441 1442 1444 1445 1446 1447 1448 1449 1453 1456 1458 1460 1461 1463 1465 1466`
	`[103] 1469 1470 1474 1475 1476 1480 1481 1482 1483 1484 1485 1486 1487 1488 1489 1490 1491`
	`[120] 1492 1494 1495 1497 1498 1499 1500 1502 1512 1513 1516 1517 1522 1523 1524 1527 1528`
	`[137] 1529 1530 1532 1533 1534 1535 1536 1538 1541 1542 1543 1544 1545 1546 1547 1548 1549`
	`[154] 1550 1552 1553 1554 1556 1558 1559 1560 1562 1566 1567 1568 1569 1571 1573 1574 1575`
	`[171] 1578 1579 1580 1581 1585 1586 1587 1588 1589 1590 1592 1593 1594 1595 1596 1597 1598`
	`[188] 1599 1600 1603 1610 1611 1612 1613 1615 1616 1617 1618 1619 1621 1624 1627 1628 1631`
	`[205] 1632 1633 1634 1636 1638 1640 1642 1643 1644 1645 1647 1650 1651 1652 1653 1655 1656`
	`[222] 1658 1662 1664 1668 1670 1671 1672 1673 1674 1677 1680 1683 1684 1687 1690 1691 1693`
	`[239] 1695 1700 1703 1705 1706 1710 1711 1713 1714 1715 1716 1717 1718 1719 1720 1721 1722`
	`[256] 1726 1728 1730 1732 1733 1734 1735 1736 1739 1740 1741 1742 1743 1744 1745 1750 1754`
	`[273] 1755 1757 1758 1762 1763 1767 1768 1769 1770 1771 1773 1774 1775 1776 1777 1779 1780`
	`[290] 1781 1783 1784 1786 1787 1788 1789 1790 1791 1792 1793 1794 1795 1799 1800 1801 1802`
	`[307] 1803 1804 1806 1808 1810 1811 1812 1814 1816 1817 1820 1823 1825 1827 1829 1830 1831`
	`[324] 1834 1835 1836 1839 1840 1841 1844 1845 1846 1848 1849 1850 1851 1852 1853 1855 1856`
	`[341] 1857 1858 1859 1860 1862 1864 1865 1866 1868 1869 1870 1871 1873 1875 1876 1877 1878`
	`[358] 1879 1880 1881 1882 1884 1885 1888 1892 1894 1895 1896 1898 1899 1900 1901 1902 1905`
	`[375] 1906 1907 1908 1909 1910 1911 1915 1916 1919 1920 1921 1922 1923 1924 1926 1927 1928`
	`[392] 1929 1931 1932 1934 1937 1938 1942 1943 1944 1945 1946 1948 1951 1953 1954 1955 1956`
	`[409] 1959 1960 1961 1963 1964 1965 1970 1971 1972 1973 1975 1977 1978 1979 1980 1985 1987`
	`[426] 1988 1990 1992 1993 1994 1995 1996 1998 2002 2003 2004 2005 2008 2009 2011 2014 2015`
	`[443] 2016 2018 2021 2022 2024 2030 2034 2036 2039 2040 2042 2044 2046 2047 2048 2049 2053`
	`[460] 2055 2056 2058 2060 2061 2063 2065 2067 2068 2070 2071 2073 2074 2075 2076 2077 2080`
	`[477] 2081 2082 2084 2085 2086 2088 2089 2090 2091 2093 2094 2095 2096 2100 2101 2103 2105`
	`[494] 2108 2110 2111 2112 2113 2114 2115 2116 2117 2118 2119 2120 2122 2123 2125 2126 2127`
	`[511] 2130 2131 2132 2134 2142 2143 2145 2146 2148 2149 2152 2153 2154 2155 2156 2159 2161`
	`[528] 2162 2164 2165 2166 2169 2170 2172 2173 2177 2178 2179 2180 2181 2183 2187 2188 2189`
	`[545] 2190 2192 2194 2195 2201 2202 2203 2204 2205 2206 2207 2210 2211 2213 2214 2215 2217`
	`[562] 2218 2219 2221 2222 2223 2225 2226 2229 2230 2231 2232 2235 2236 2237 2238 2240 2241`
	`[579] 2242 2243 2245 2246 2247 2248 2250 2251 2252 2253 2254 2255 2256 2257 2259 2260 2261`
	`[596] 2262 2264 2265 2266 2269 2271 2272 2273 2274 2275 2276 2277 2280 2282 2283 2286 2290`
	`[613] 2293 2294 2295 2296 2297 2298 2299 2302 2303 2304 2305 2308 2309 2310 2311 2312 2314`
	`[630] 2315 2317 2318 2319 2320 2321 2322 2323 2324 2325 2326 2329 2330 2331 2332 2333 2334`
	`[647] 2335 2337 2339 2341 2342 2343 2344 2347 2348 2349 2350 2351 2352 2354 2355 2356 2357`
	`[664] 2358 2359 2360 2362 2363 2364 2365 2366 2367 2368 2370 2371 2372 2373 2374 2376 2377`
	`[681] 2378 2379 2380 2381 2382 2384 2385 2386 2387 2388 2390 2391 2392 2393 2394 2395 2396`
	`[698] 2397 2399 2400 2401 2403 2406 2407 2408 2409 2410 2411 2412 2413 2415 2416 2417 2418`
	`[715] 2419 2420 2421 2422 2423 2424 2425 2427 2428 2430 2431 2432 2433 2434 2436 2438 2440`
	`[732] 2441 2442 2443 2444 2446 2447 2449 2450 2451 2452 2453 2454 2455 2456 2457 2458 2461`
	`[749] 2462 2464 2466 2467 2468 2469 2470 2471 2472 2473 2474 2475 2476 2477 2478 2479 2481`
	`[766] 2482 2483 2484 2485 2486 2487 2488 2489 2491 2492 2494 2495 2496 2497 2499 2500 2501`
	`[783] 2504 2505 2506 2507 2508 2509 2510 2511 2514 2515 2516 2517 2518 2520 2522 2523 2524`
	`[800] 2525 2526 2527 2528 2529 2530 2531 2532 2533 2534 2535 2536 2537 2539 2540 2541 2542`
	`[817] 2543 2544 2545 2546 2548 2550 2552 2553 2554 2555 2556 2558 2559 2561 2562 2563 2564`
	`[834] 2565 2567 2568 2569 2570 2571 2572 2574 2575 2576 2577 2578 2579 2580 2581 2582 2583`
	`[851] 2584 2585 2586 2588 2589 2590 2591 2592 2593 2595 2597 2598 2599 2601 2602 2604 2605`
	`[868] 2606 2608 2609 2611 2612 2613 2614 2615 2616 2617 2618 2619 2620 2622 2623 2624 2625`
	`[885] 2626 2627 2630 2631 2632 2633 2634 2635 2636 2637 2638 2639 2640 2641 2643 2644 2645`
	`[902] 2646 2647 2648 2650 2651 2652 2653 2654 2657 2660 2663 2664 2665 2666 2667 2668 2669`
	`[919] 2670 2671 2672 2673 2674 2675 2676 2678 2679 2680 2681 2683 2685 2687 2689 2690 2691`
	`[936] 2692 2693 2694 2696 2697 2698 2699 2700 2701 2702 2703 2704 2705 2706 2708 2709 2710`
	`[953] 2711 2712 2714 2715 2716 2718 2719 2720 2721 2722 2723 2724 2726 2727 2728 2730 2731`
	`[970] 2732 2733 2734 2735 2736 2737 2738 2739 2740 2741 2742 2743 2744 2746 2747 2748 2749`
	`[987] 2750 2751 2752 2753 2755 2756 2757 2758 2759 2760 2761 2762 2764 2766`
	`[ reached getOption("max.print") -- omitted 5284 entries ]`

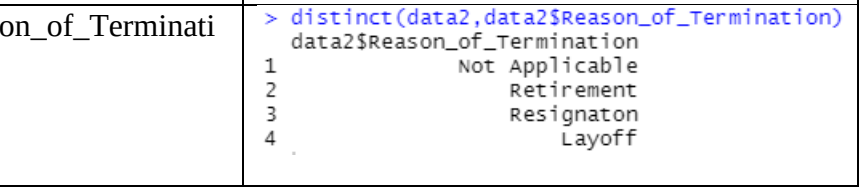
unique(data2$Age)	`> unique(data2$Age)` `[1] 52 53 54 55 56 57 58 59 60 61 49 50 51 47 48 44 45 46 42 43 39 40 41 62 63 64 65 38 37 36 35 34`
unique(data2$Year_of_Service)	`> unique(data2$Year_of_Service)` `[1] 17 18 19 20 21 22 23 24 25 26 16 15 14 13 12 11 10 9 8 7`

unique(data2$City)	`> unique(data2$City)`
	`[1] "Vancouver" "Terrace" "Nanaimo" "Nelson" "Kelowna"`
	`[6] "Victoria" "Kamloops" "Fort St John" "Surrey" "Vernon"`
	`[11] "Quesnel" "Chilliwack" "Dawson Creek" "Squamish" "New Westminster"`
	`[16] "Port Coquitlam" "Cortes Island" "Burnaby" "Bella Bella" "Cranbrook"`
	`[21] "Williams Lake" "Trail" "Prince George" "Richmond" "Grand Forks"`
	`[26] "West Vancouver" "Abbotsford" "Aldergrove" "Langley" "North Vancouver"`
	`[31] "White Rock" "Fort Nelson" "Haney" "Valemount" "Ocean Falls"`
	`[36] "Pitt Meadows" "Princeton" "Dease Lake" "Blue River"`

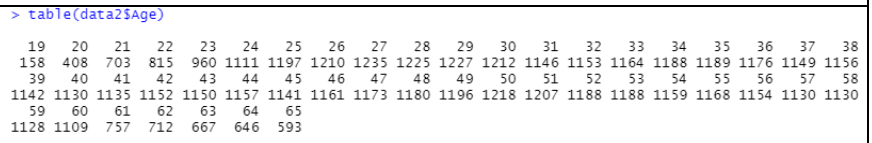
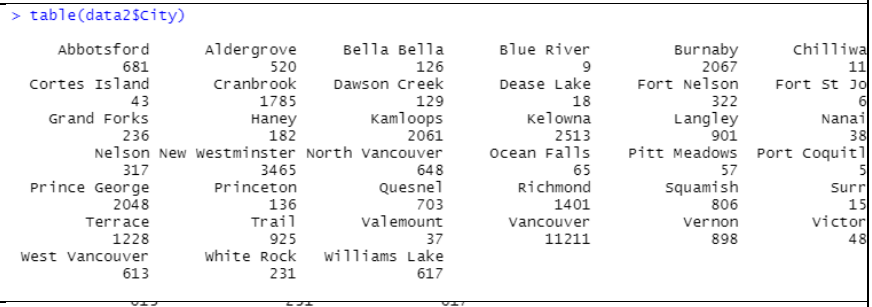
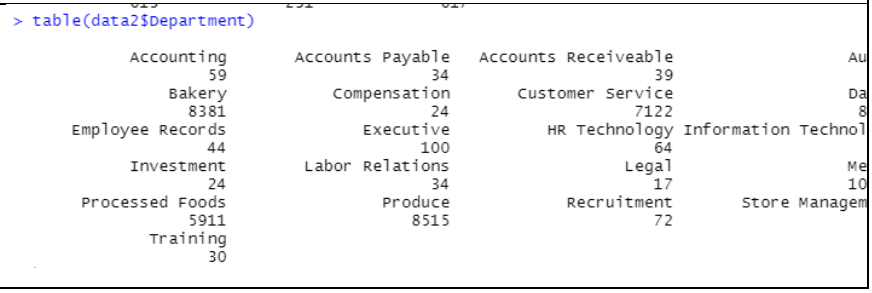
unique(data2$Department)	`> unique(data2$Department)`
	`[1] "Executive" "Store Management" "Meats"`
	`[5] "Training" "Labor Relations" "HR Technology"`
	`[9] "Compensation" "Legal" "Produce"`
	`[13] "Bakery" "Information Technology" "Accounts Payable"`
	`[17] "Accounting" "Investment" "Dairy"`
	`[21] "Customer Service"`

unique(data2$Job)	
unique(data2$Store_Code)	
unique(data2$Type_of_Termination)	
unique(data2$Year_of_Status)	
unique(data2$Business_Unit)	

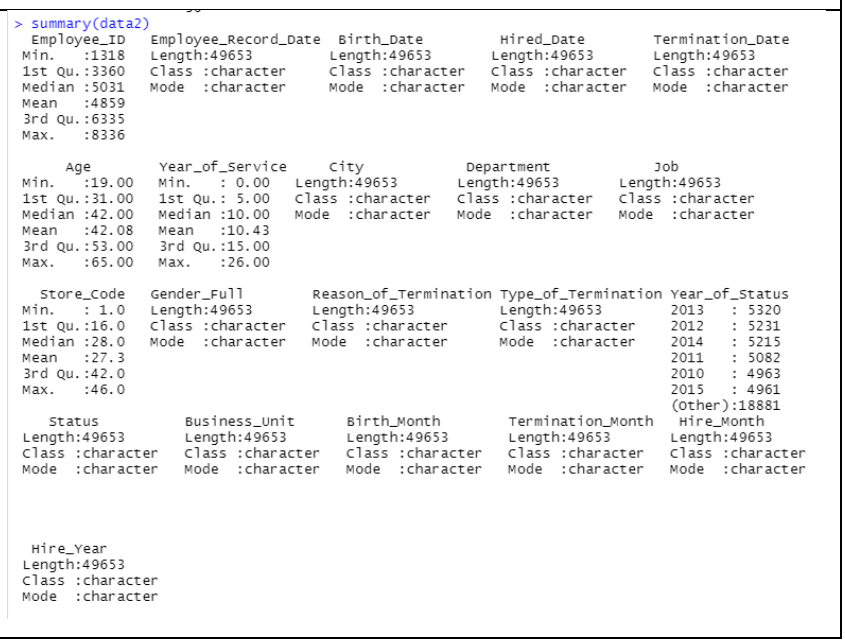
These commands is to show all unique data inside an attribute.

Source Code	**Output**
distinct(data2,data2$Reason_of_Termination)	

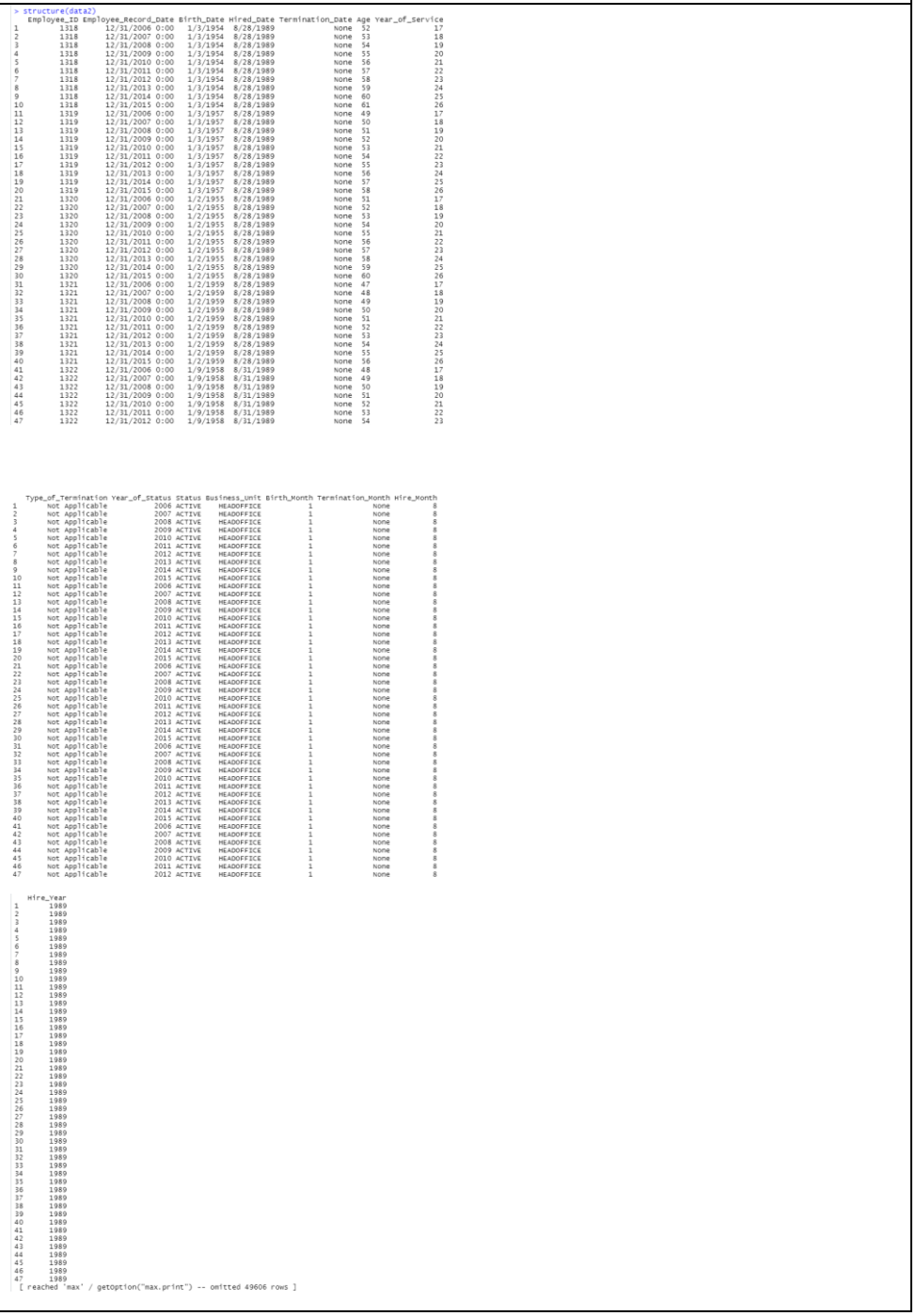
The distinct command is to help identify the types of answer given in a dataset.

Source Code	Output
table(data2$Age)	
table(data2$Gender_Full)	
table(data2$City)	
table(data2$Department)	

The table command is to catergorize the data to display the unique category and the frequency of the data.

Source Code	Output
summary(data2)	

The summary command is to show the summary of the dataset imported, which include the minimum, first quartile, median, mean their quartile and maximum.

Source Code	Output
structure(data2)	

```
> structure(data2)
   Employee_ID Employee_Record_Date Birth_Date Hired_Date Termination_Date Age Year_of_Service
1         1318      12/31/2006 0:00   1/3/1954  8/28/1989             None  52              17
2         1318      12/31/2007 0:00   1/3/1954  8/28/1989             None  53              18
3         1318      12/31/2008 0:00   1/3/1954  8/28/1989             None  54              19
4         1318      12/31/2009 0:00   1/3/1954  8/28/1989             None  55              20
5         1318      12/31/2010 0:00   1/3/1954  8/28/1989             None  56              21
6         1318      12/31/2011 0:00   1/3/1954  8/28/1989             None  57              22
7         1318      12/31/2012 0:00   1/3/1954  8/28/1989             None  58              23
8         1318      12/31/2013 0:00   1/3/1954  8/28/1989             None  59              24
9         1318      12/31/2014 0:00   1/3/1954  8/28/1989             None  60              25
10        1318      12/31/2015 0:00   1/3/1954  8/28/1989             None  61              26
11        1319      12/31/2006 0:00   1/3/1957  8/28/1989             None  49              17
12        1319      12/31/2007 0:00   1/3/1957  8/28/1989             None  50              18
13        1319      12/31/2008 0:00   1/3/1957  8/28/1989             None  51              19
14        1319      12/31/2009 0:00   1/3/1957  8/28/1989             None  52              20
15        1319      12/31/2010 0:00   1/3/1957  8/28/1989             None  53              21
16        1319      12/31/2011 0:00   1/3/1957  8/28/1989             None  54              22
17        1319      12/31/2012 0:00   1/3/1957  8/28/1989             None  55              23
18        1319      12/31/2013 0:00   1/3/1957  8/28/1989             None  56              24
19        1319      12/31/2014 0:00   1/3/1957  8/28/1989             None  57              25
20        1319      12/31/2015 0:00   1/3/1957  8/28/1989             None  58              26
21        1320      12/31/2006 0:00   1/2/1955  8/28/1989             None  51              17
22        1320      12/31/2007 0:00   1/2/1955  8/28/1989             None  52              18
23        1320      12/31/2008 0:00   1/2/1955  8/28/1989             None  53              19
24        1320      12/31/2009 0:00   1/2/1955  8/28/1989             None  54              20
25        1320      12/31/2010 0:00   1/2/1955  8/28/1989             None  55              21
26        1320      12/31/2011 0:00   1/2/1955  8/28/1989             None  56              22
27        1320      12/31/2012 0:00   1/2/1955  8/28/1989             None  57              23
28        1320      12/31/2013 0:00   1/2/1955  8/28/1989             None  58              24
29        1320      12/31/2014 0:00   1/2/1955  8/28/1989             None  59              25
30        1320      12/31/2015 0:00   1/2/1955  8/28/1989             None  60              26
31        1321      12/31/2006 0:00   1/2/1959  8/28/1989             None  47              17
32        1321      12/31/2007 0:00   1/2/1959  8/28/1989             None  48              18
33        1321      12/31/2008 0:00   1/2/1959  8/28/1989             None  49              19
34        1321      12/31/2009 0:00   1/2/1959  8/28/1989             None  50              20
35        1321      12/31/2010 0:00   1/2/1959  8/28/1989             None  51              21
36        1321      12/31/2011 0:00   1/2/1959  8/28/1989             None  52              22
37        1321      12/31/2012 0:00   1/2/1959  8/28/1989             None  53              23
38        1321      12/31/2013 0:00   1/2/1959  8/28/1989             None  54              24
39        1321      12/31/2014 0:00   1/2/1959  8/28/1989             None  55              25
40        1321      12/31/2015 0:00   1/2/1959  8/28/1989             None  56              26
41        1322      12/31/2006 0:00   1/9/1958  8/31/1989             None  48              17
42        1322      12/31/2007 0:00   1/9/1958  8/31/1989             None  49              18
43        1322      12/31/2008 0:00   1/9/1958  8/31/1989             None  50              19
44        1322      12/31/2009 0:00   1/9/1958  8/31/1989             None  51              20
45        1322      12/31/2010 0:00   1/9/1958  8/31/1989             None  52              21
46        1322      12/31/2011 0:00   1/9/1958  8/31/1989             None  53              22
47        1322      12/31/2012 0:00   1/9/1958  8/31/1989             None  54              23

   Type_of_Termination Year_of_Status Status Business_Unit Birth_Month Termination_Month Hire_Month
1       Not Applicable           2006 ACTIVE    HEADOFFICE           1              None          8
2       Not Applicable           2007 ACTIVE    HEADOFFICE           1              None          8
3       Not Applicable           2008 ACTIVE    HEADOFFICE           1              None          8
4       Not Applicable           2009 ACTIVE    HEADOFFICE           1              None          8
5       Not Applicable           2010 ACTIVE    HEADOFFICE           1              None          8
6       Not Applicable           2011 ACTIVE    HEADOFFICE           1              None          8
7       Not Applicable           2012 ACTIVE    HEADOFFICE           1              None          8
8       Not Applicable           2013 ACTIVE    HEADOFFICE           1              None          8
9       Not Applicable           2014 ACTIVE    HEADOFFICE           1              None          8
10      Not Applicable           2015 ACTIVE    HEADOFFICE           1              None          8
11      Not Applicable           2006 ACTIVE    HEADOFFICE           1              None          8
12      Not Applicable           2007 ACTIVE    HEADOFFICE           1              None          8
13      Not Applicable           2008 ACTIVE    HEADOFFICE           1              None          8
14      Not Applicable           2009 ACTIVE    HEADOFFICE           1              None          8
15      Not Applicable           2010 ACTIVE    HEADOFFICE           1              None          8
16      Not Applicable           2011 ACTIVE    HEADOFFICE           1              None          8
17      Not Applicable           2012 ACTIVE    HEADOFFICE           1              None          8
18      Not Applicable           2013 ACTIVE    HEADOFFICE           1              None          8
19      Not Applicable           2014 ACTIVE    HEADOFFICE           1              None          8
20      Not Applicable           2015 ACTIVE    HEADOFFICE           1              None          8
21      Not Applicable           2006 ACTIVE    HEADOFFICE           1              None          8
22      Not Applicable           2007 ACTIVE    HEADOFFICE           1              None          8
23      Not Applicable           2008 ACTIVE    HEADOFFICE           1              None          8
24      Not Applicable           2009 ACTIVE    HEADOFFICE           1              None          8
25      Not Applicable           2010 ACTIVE    HEADOFFICE           1              None          8
26      Not Applicable           2011 ACTIVE    HEADOFFICE           1              None          8
27      Not Applicable           2012 ACTIVE    HEADOFFICE           1              None          8
28      Not Applicable           2013 ACTIVE    HEADOFFICE           1              None          8
29      Not Applicable           2014 ACTIVE    HEADOFFICE           1              None          8
30      Not Applicable           2015 ACTIVE    HEADOFFICE           1              None          8
31      Not Applicable           2006 ACTIVE    HEADOFFICE           1              None          8
32      Not Applicable           2007 ACTIVE    HEADOFFICE           1              None          8
33      Not Applicable           2008 ACTIVE    HEADOFFICE           1              None          8
34      Not Applicable           2009 ACTIVE    HEADOFFICE           1              None          8
35      Not Applicable           2010 ACTIVE    HEADOFFICE           1              None          8
36      Not Applicable           2011 ACTIVE    HEADOFFICE           1              None          8
37      Not Applicable           2012 ACTIVE    HEADOFFICE           1              None          8
38      Not Applicable           2013 ACTIVE    HEADOFFICE           1              None          8
39      Not Applicable           2014 ACTIVE    HEADOFFICE           1              None          8
40      Not Applicable           2015 ACTIVE    HEADOFFICE           1              None          8
41      Not Applicable           2006 ACTIVE    HEADOFFICE           1              None          8
42      Not Applicable           2007 ACTIVE    HEADOFFICE           1              None          8
43      Not Applicable           2008 ACTIVE    HEADOFFICE           1              None          8
44      Not Applicable           2009 ACTIVE    HEADOFFICE           1              None          8
45      Not Applicable           2010 ACTIVE    HEADOFFICE           1              None          8
46      Not Applicable           2011 ACTIVE    HEADOFFICE           1              None          8
47      Not Applicable           2012 ACTIVE    HEADOFFICE           1              None          8

   Hire_Year
1       1989
2       1989
3       1989
4       1989
5       1989
6       1989
7       1989
8       1989
9       1989
10      1989
11      1989
12      1989
13      1989
14      1989
15      1989
16      1989
17      1989
18      1989
19      1989
20      1989
21      1989
22      1989
23      1989
24      1989
25      1989
26      1989
27      1989
28      1989
29      1989
30      1989
31      1989
32      1989
33      1989
34      1989
35      1989
36      1989
37      1989
38      1989
39      1989
40      1989
41      1989
42      1989
43      1989
44      1989
45      1989
46      1989
47      1989
 [ reached 'max' / getOption("max.print") -- omitted 49606 rows ]
```

The structure command is to view all of the dataset.

Pre-processing Data

Data pre-processing refers to any operation performed on raw data to prepare it for subsequent processing. Historically, it served as the initial phase in the data mining process.

```
#Pre-proccessing data
avg_age2006 =colMeans(data2 %>% filter(Year_of_Status=="2006") %>% select(Age))
avg_age2007 =colMeans(data2 %>% filter(Year_of_Status=="2007") %>% select(Age))
avg_age2008 =colMeans(data2 %>% filter(Year_of_Status=="2008") %>% select(Age))
avg_age2009 =colMeans(data2 %>% filter(Year_of_Status=="2009") %>% select(Age))
avg_age2010 =colMeans(data2 %>% filter(Year_of_Status=="2010") %>% select(Age))
avg_age2011 =colMeans(data2 %>% filter(Year_of_Status=="2011") %>% select(Age))
avg_age2012 =colMeans(data2 %>% filter(Year_of_Status=="2012") %>% select(Age))
avg_age2013 =colMeans(data2 %>% filter(Year_of_Status=="2013") %>% select(Age))
avg_age2014 =colMeans(data2 %>% filter(Year_of_Status=="2014") %>% select(Age))
avg_age2015 =colMeans(data2 %>% filter(Year_of_Status=="2015") %>% select(Age))
```

The command above is to assign a value to a variable which are than used for later coding.

Analysis 1: Observation on the changes in between the year 2006 to 2015

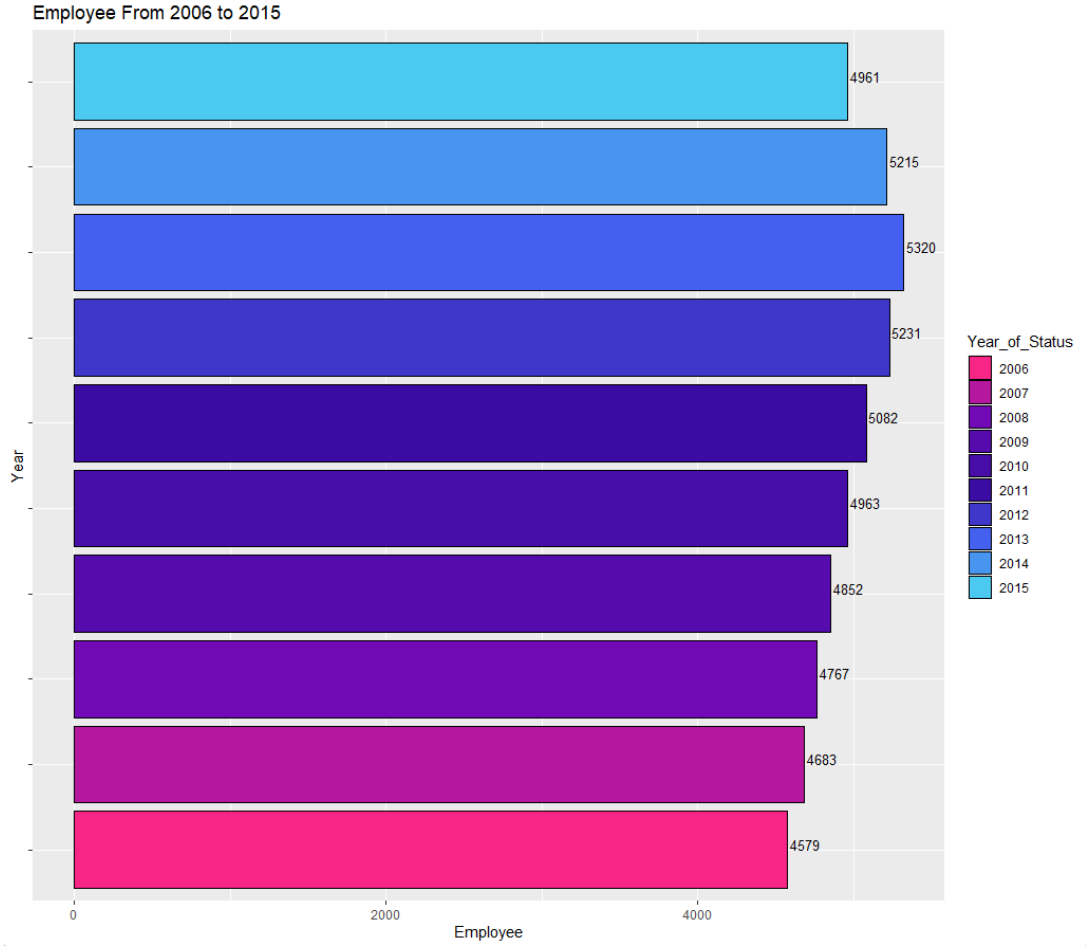
Analysis 1.1: Observations on the changes in the number of employees between the year 2006 to 2015

Source code

```
#Number of employee
data2 %>% group_by(Year_of_Status) %>% summarise(Employee=n()) %>%
  ggplot(aes(x=Year_of_Status,y=Employee,fill=Year_of_Status))+
  geom_bar(col="black", stat = "identity",position = "dodge")+
  ggtitle("Employee From 2006 to 2015")+coord_flip()+xlab("Year")+
  geom_text(aes(label=Employee),hjust=-0.1,vjust=0,col="black",size=3.5)+
  theme(axis.text.y = element_blank())+
  scale_fill_manual(values =c('#f72585','#b5179e','#7209b7','#560bad','#480ca8',
                              '#3a0ca3','#3f37c9','#4361ee','#4895ef','#4cc9f0'))
```

Line	Explanation
1	Catergorizing summary of year of status using the group by function
2-7	Generating a bar chart
3	Adding titles to the bar chart
7	Adding colour labels for different variable in the bar chart

Data Visualization



Observation:

Through the data visualization above, we could see that 2013 has the most employees and the year which has the least employees is 2006. Other than this, we could see a steady increase of employees joining the workforce from year 2006 to year 2013. But after 2013, there is a constant decrease in the amount of employees in the company.

This shows that the company has been steadily employing new employees into the company and employee retention is stable throughout the year but after the year 2015, employees have been leaving and the company has also decreased the amount of hiring.

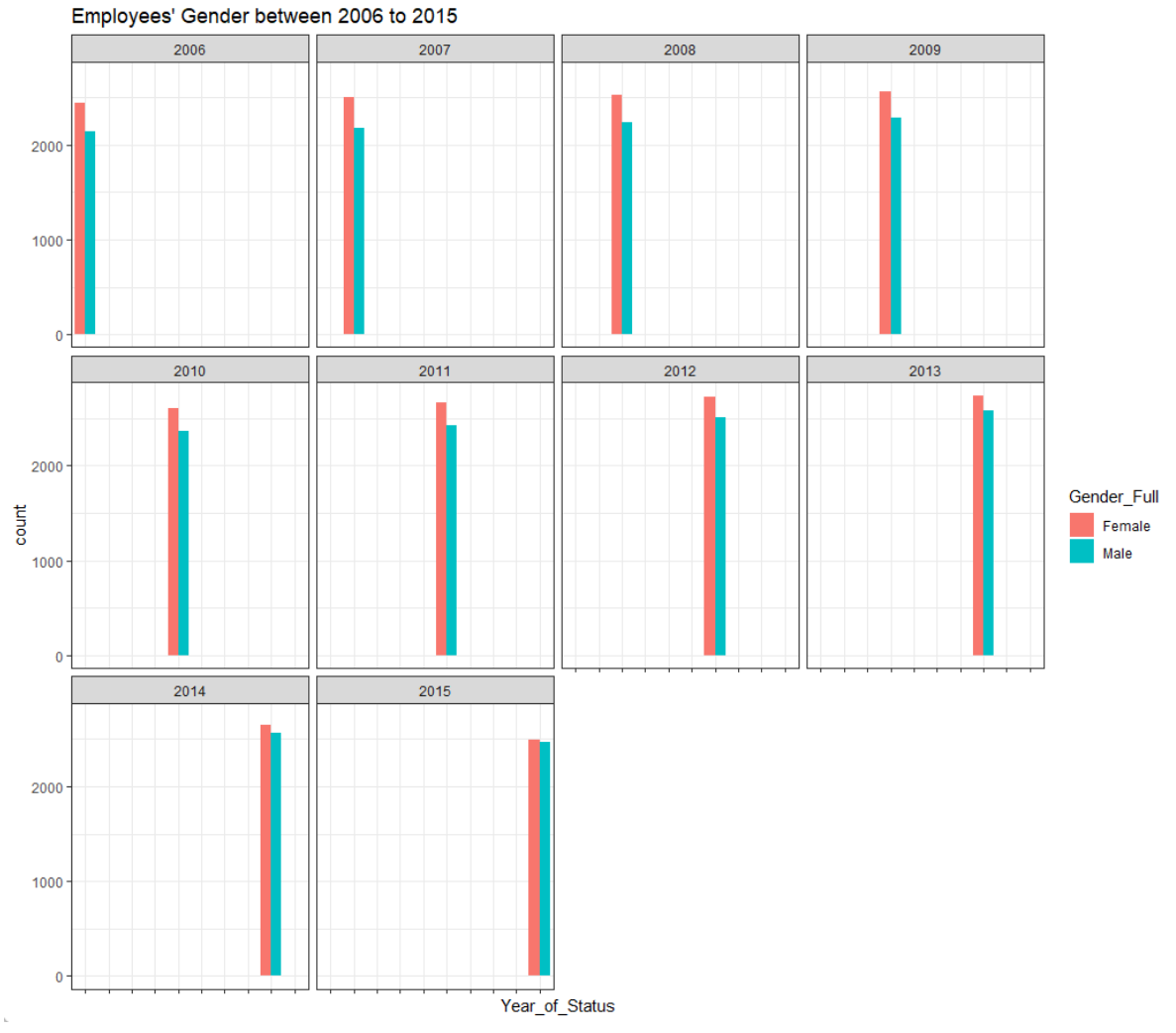
Analysis 1.2: Observations on the changes in the gender of employees between the year 2006 to 2015

Source code

```
#Gender
ggplot(data2,aes(x=Year_of_Status,fill=Gender_Full))+geom_bar(position="dodge")+
  facet_wrap(~Year_of_Status)+theme_bw()+ggtitle("Employees' Gender between 2006 to 2015")+
  theme(axis.text.x = element_blank())
```

line	Explanation
1-3	Generating bar chart
2	1. Creating multiple panel 2. Creating the title for the bar chart

Data Visualization



Observation:

Based on the data visualization above, we could see that the gender ratio between male and female have been changing toward a pattern along the year. Starting from the year 2006, we could see that the company has added more female employees into the workforce, but throughout the year, more male have also joined the company.

This shows that during the year 2006, more females are interested in joining the company, but as the year has passed, more male are also interested in joining the company.

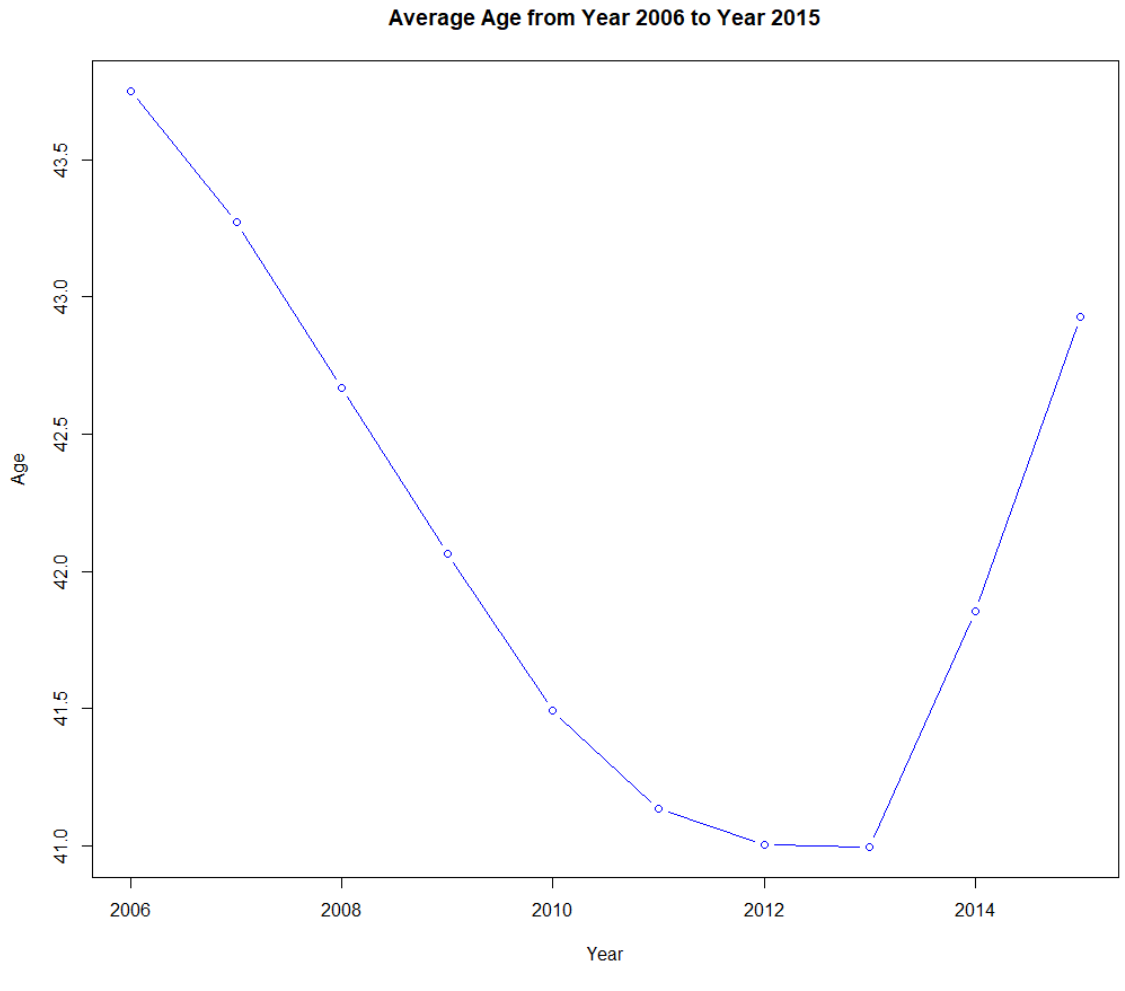
Analysis 1.3: Observations on the changes in the average age group of employees between the year 2006 to 2015

Source code

```
#Average Age Group
Year = c("2006","2007","2008","2009","2010","2011","2012","2013","2014","2015")

avg_ageG = cbind(avg_age2006,avg_age2007,avg_age2008,avg_age2009,avg_age2010,avg_age2011,avg_age2012,avg_age2013,avg_age2014,avg_age2015)

plot(Year,avg_ageG,type="b",main="Average Age from Year 2006 to Year 2015",ylab = "Age",col="blue")
```

line	Explanation
1	Assigning the variable year with a vector of year 2006 - 2015
2	Assign varaible avg_ageG with the average age from 2006-2015 using the cbind funciton
3	Generate the line chart

Data Visualization



Observation:

Based on the data visualization shown above, we could see that the average age has been decreasing along the year. Starting from 2006, to 2012, the company has seen a steady decline of average age in their company but in the year 2014, there has been an increase in average age in the company.

This shows that, more and more old employees are leaving the workforce and more and more young employees are hired to fill in the empty spots in the company, but in 2014, there is a spike of average age, this shows us that the company has turned toward more older employees to help fill in the workforce.

Analysis 1.4: Observations on the changes in the city of residing of employees between the year 2006 to 2015

Source code

```
#City
ggplot(data2,aes(x=Year_of_Status))+geom_bar(position="dodge",fill="#ffa7a6")+facet_wrap(~City)+theme_bw()+
  ggtitle("Employees' City between 2006 to 2015")+xlab("Year (2006 - 2015)")+ylab("Number of Employee")+
  theme(axis.text.x = element_blank())
```

line	Explanation
1-3	Generate the bar chart
1	- Summarize the data by the year of status - Create multiple panel to showcase different cities
2	Create title and label for the bar chart

Data Visualization

Observation:

Based on the data visualization shown above, we could see that most of the company is from Vancouver. But cities such as Victoria, Kamloops, Kelowna, Nanaimo, New Westminster,Cranbrook and Burnaby also have a decent amount of employees working in it. Cities like Valemount, Pitt Meadows, Ocean Falls and Blue Rivers have fewer employees residing.

This shows that the company is mostly centered in Vancouver. The amount of employees from Vancouver is double the amount compared to other places. The company also has a decent presence in cities like Victoria, Kamloops, Kelowna, Nanaimo, New Westminster, Cranbrooks and Burnby, that's why we could see a decent amount of employees coming from these cities.

But with proper observation, we could see that all the graph have something in common which is the graph will see a small increase throughout the year but at the end of each graph there is a small decline, I believe this is caused by policy changes in the country as only so will it affect the hiring rate of the company.

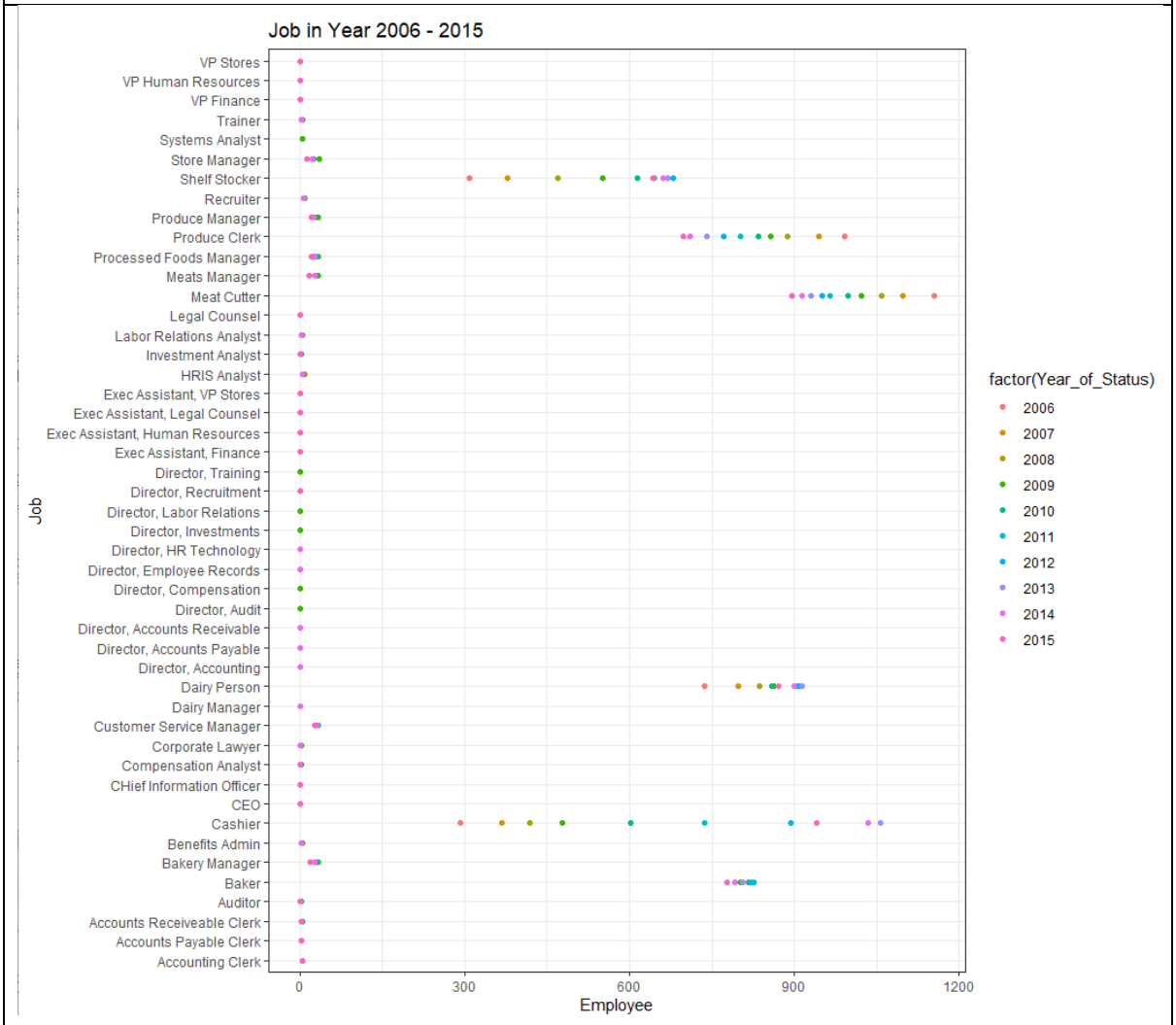
Analysis 1.5: Observations on the changes of the job of employees between the year 2006 to 2015

Source code

```
#Job
data2 %>% group_by(Year_of_Status,Job) %>% select(Job) %>% summarise(Employee = n()) %>%
  ggplot(aes(y=Job,group=Year_of_Status,color=factor(Year_of_Status),x=Employee))+geom_point(size=1.5)+
  ggtitle("Job in Year 2006 - 2015")+theme_bw()
```

line	Explanation
1-3	Generating scatter plot grapgh
1	Summarizing the dataset by grouping the year of status according to job
3	Creating title and label for the scatter plot graph

Data Visualization



Observation:

Based on the data visualization above, we could observe that the company hires a lot of cashiers, baker, dairy person, meat cutter, produce clerk and shelf stocker. Other positions like bakery manager, customer service manager, meat manager, processed food manager, produce manager and store manager are also needed to operate the business.

This shows that the company require lots of low level employee like cashier, baker, dairy person, meat cutter, produce clerk and shelf stocker to conduct their daily business whereas position like bakery manager, customer service manager, meat manager, processed food manager, produce manager and store manager are needed to supervise them.

We could also see that the amount of low level employees like cashier, baker, dairy person, meat cutter, produce clerk and shelf stocker varies throughout the year 2006 to 2015. Which shows the hiring interest of the company.

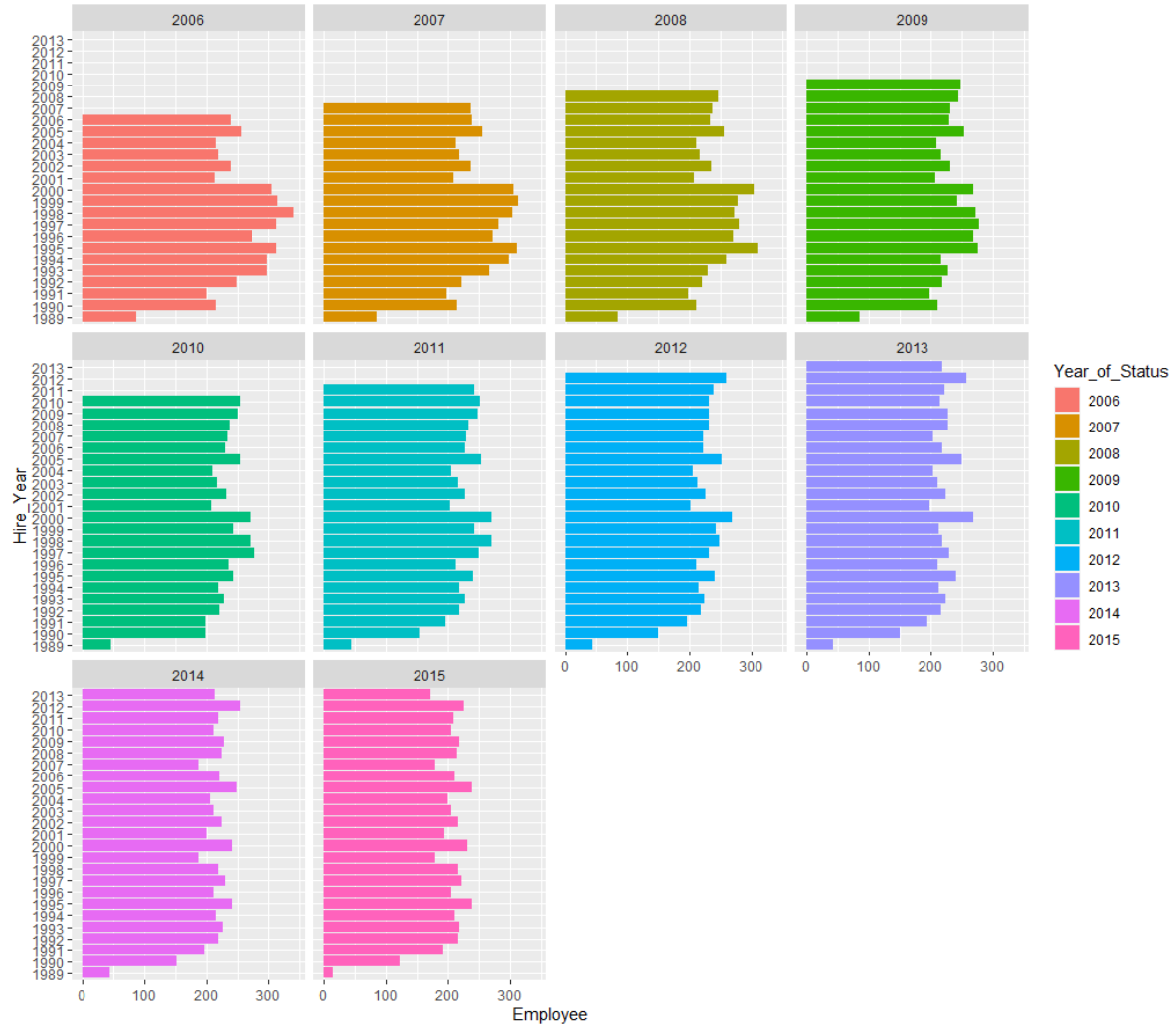
Analysis 1.6: Observations on the changes in the number of employees between the year 2006 to 2015

Source code

```
#Hire ( It show how many new employee hired and the number of employee reduced year by year)
data2 %>% group_by(Year_of_Status,Hire_Year) %>% select(Hire_Year,Year_of_Status) %>% summarise(Employee = n()) %>%
  ggplot(aes(x=Hire_Year,y=Employee,fill = Year_of_Status))+geom_bar(stat="identity")+coord_flip()+
  facet_wrap(~Year_of_Status)
```

line	Explanation
1	Summarize the data by grouping the year of status according to the hire year while selecting hire year according to the year of status
2	Generating bar chart
3	Creating multiple panels to show different data

Data Visualization



Observation:

Based on the data visualization above, we could see that the company stopped hiring people in 2013. The second observation is that throughout the year, the company has seen a decline in the amount of employees who stayed. Employees from the year 1995 to 1989 have been decreasing at a huge rate and employees that join during 2010 to 2015 are found to decrease along the year.

This shows that the young employees have less interest in working with the company whereas a lot of older employees are leaving the company due to retirement issues. Something might have happened during the year 2014 and 2015 which led to the company stop hiring new employees to join the workforce. It also shows that the company has a hard time retaining new young aged employees which lead to high termination rate among newly joined employees.

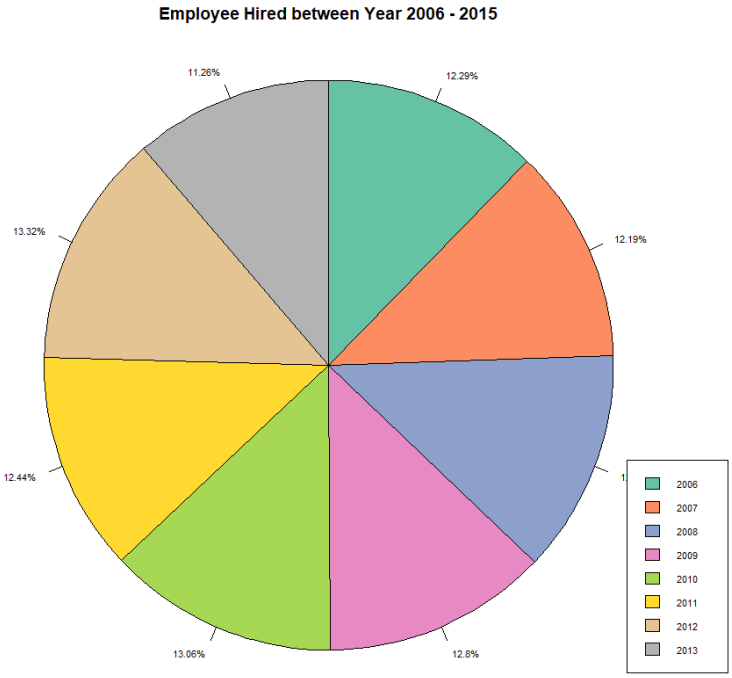
Analysis 1.7: Observations on the changes in the hiring rate of employees between the year 2006 to 2015

Source code

```
#hire percentage
hire_em_peryear = rbind(nrow(data2 %>% filter(Year_of_Status=="2006",Hire_Year=="2006") %>% select(Age)),
                        nrow(data2 %>% filter(Year_of_Status=="2007",Hire_Year=="2007") %>% select(Age)),
                        nrow(data2 %>% filter(Year_of_Status=="2008",Hire_Year=="2008") %>% select(Age)),
                        nrow(data2 %>% filter(Year_of_Status=="2009",Hire_Year=="2009") %>% select(Age)),
                        nrow(data2 %>% filter(Year_of_Status=="2010",Hire_Year=="2010") %>% select(Age)),
                        nrow(data2 %>% filter(Year_of_Status=="2011",Hire_Year=="2011") %>% select(Age)),
                        nrow(data2 %>% filter(Year_of_Status=="2012",Hire_Year=="2012") %>% select(Age)),
                        nrow(data2 %>% filter(Year_of_Status=="2013",Hire_Year=="2013") %>% select(Age)))

hire_en_peryear_percent <-paste0(round(hire_em_peryear/sum(hire_em_peryear)*100,2),"%")

colour = brewer.pal(length(hire_em_peryear),"Set2")

pie(hire_em_peryear,labels=hire_en_peryear_percent,cex=0.7,radius=1,main="Employee Hired between Year 2006 - 2015",border="black",
    clockwise = TRUE,col=colour)
legend("bottomright",c("2006","2007","2008","2009","2010","2011","2012","2013"),cex=0.7,
       fill=colour)
```

line	Explanation
1	Variable name “hire_em_peryear” is assigned the attribute year of status and hire year with the filter of each year while selecting employees age which is in the year
2	Store the turned data into percentage into a variable called hire_en_peryear_percent
3	Assign a set of colour tinto the variable colour
4	Generate a pie chart
5	Generate a legend to state information of the pie chart

Data Visualization



Observation:

Based on the visualization above, we could see that the company have a constant rate of hiring throughout the year 2006 to 2013 with the highest going to 2012 with the hiring rate of 13.2% and the lowest which is 2013 with the percentage of 11.2%

This shows that the company has constantly hired employees throughout the year without any drastic action. It also shows that the public have no issues with working with the company as there is no decline in the hiring rate of the company. But something might have happened during year 2014 and 2015 which caused the company to stop hiring new employee

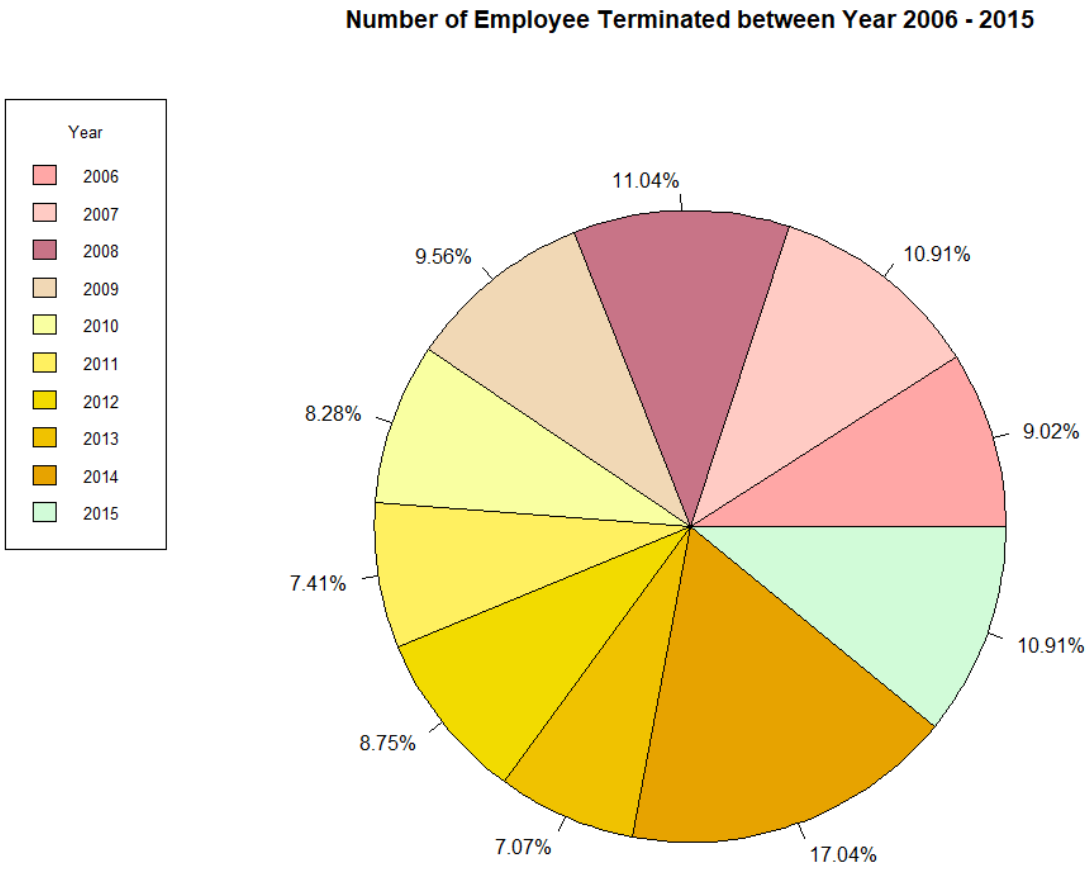
Analysis 1.8: Observations on the changes in the number of terminated employees between the year 2006 to 2015

Source code

```
#Number of terminate
terminate = table(data2 %>% filter(Status == "TERMINATED") %>% group_by(Year_of_Status) %>% select(Year_of_Status))
p_terminate = paste0(round(terminate/sum(terminate)*100,2),"%")
pie(terminate,label=p_terminate,main=("Number of Employee Terminated between Year 2006 - 2015"),
    col=c("#ffa7a6","#ffcbc4","#c87487","#f1d8b5","#f9ffa1","#fff060","#f2db00","#f0c200","#e7a300","#d1fbd8"))
legend("topleft",c("2006","2007","2008","2009","2010","2011","2012","2013","2014","2015"),cex = 0.8,
       fill = c("#ffa7a6","#ffcbc4","#c87487","#f1d8b5","#f9ffa1","#fff060","#f2db00","#f0c200","#e7a300","#d1fbd8"),
       title="Year")
```

line	Explanation
1	The variable terminate is used to select and store status which filter employees which is terminated using the table function while group by the year of status while selecting the year of status
2	The variable name p_terminate is used to turn the data into percentage format
3	Generate a pie chart
4	Generate a legend to state the information of the pie chart

Data Visualization



Observation:

Based on the visualization above, we could observe that the company has terminated a lot of its employees at the year 2014 with the percentage of 17.04% of the total employee terminated throughout the year 2006 to year 2015. The year with the second highest termination rate is the year 2008 with the termination rate of 11.04% of the total employees the company has terminated throughout the year 2006 to the year 2015.

This shows that something has happened during these three years respectively, specifically 2014, year 2008 and year 2015. One of the major reasons why employees leave the company is due to company policy or local policy changes which plays a huge factor.

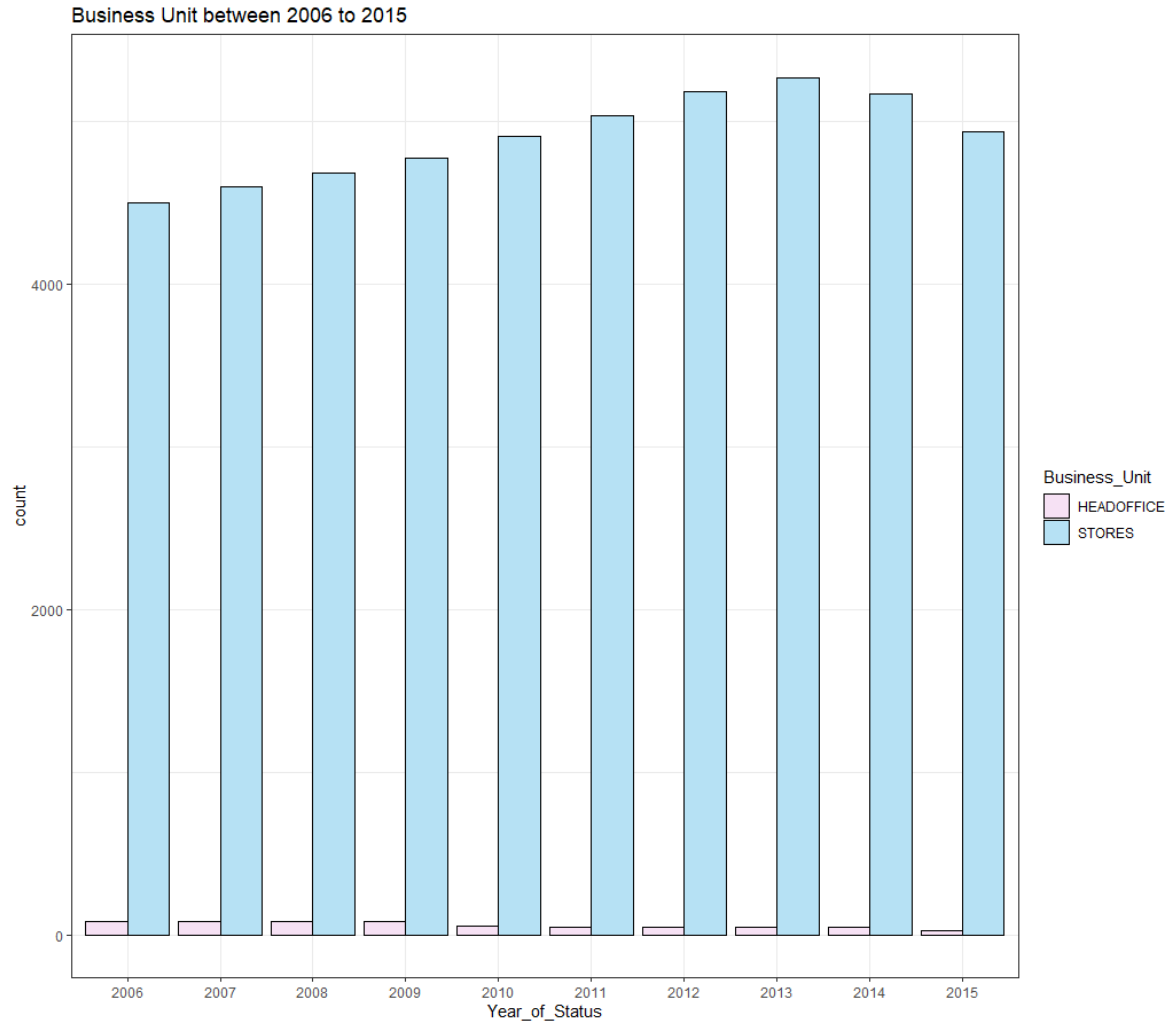
Analysis 1.9: Observations on the changes in the employees' business unit between the year 2006 to 2015

Source code

```
#Business Unit
ggplot(data2,aes(x=Year_of_Status,fill=Business_Unit))+geom_bar(position="dodge",col="black")+
  theme_bw()+ggtitle("Business Unit between 2006 to 2015")+scale_fill_manual(values=c("#f6e1f4","#b6e1f4"))
```

line	Explanation
1-2	Generating the bar chart
2	Creating the title and the label for the bar chart

Data Visualization



Observation:

Based on the data visualization above, we could observe the ratio of employee who work in the headoffice as compared to employees who have worked in stores. Throughout the year, the amount of employees that are working in the headoffice have seen a steady decline, whereas the amount of employees employed to work at the stores have seen an increase throughout the year 2006 to year 2013, then a small decline at the year 2014 to year 2015.

In my opinion, I believe that the amount of employees that have worked in the stores have decreased. There is something related with the amount of employees working in the headoffice as companies should have sufficient employees to efficiently manage the stores.

Conclusion of Analysis 1: Observation on the changes between year 2006 to 2015

After conducting various analysis based on: -

- Observations on the changes in the number of employees between the year 2006 to 2015
- Observations on the changes in the gender of employees between the year 2006 to 2015
- Observations on the changes in the average age group of employees between the year 2006 to 2015
- Observations on the changes in the city of residing of employees between the year 2006 to 2015
- Observations on the changes of the job of employees between the year 2006 to 2015
- Observations on the changes in the number of employees between the year 2006 to 2015
- Observations on the changes in the hiring rate of employees between the year 2006 to 2015
- Observations on the changes in the number of terminated employees between the year 2006 to 2015
- Observations on the changes in the employees' business unit between the year 2006 to 2015

It could be concluded that: -

- The number of employees is decreasing along the years
- The company are centered at the city Vancouver.
- The balance between male and female are closing along the year
- The company have hired a lot of cashiers, baker, dairy person, meat cutter, produce clerk and shelf stocker.
- The company have hired a lot of young employees throughout the years.
- The company stop hiring at the year 2014
- Throughout the year 2006 – 2013, the company have a constant hiring rate.
- The year 2014 have the most terminated employee
- As the amount of headoffice employee decrease, the amount of store employees decreases too.

Evidence: -

It is mentioned by the Federal Reserve Bank of San Francisco, that the current unemployment rate is due to low rates of people moving into unemployment.(frbsf.org,2020) Relating back to the company, reasons why there is a spike of terminated employee at the year 2014 is due to the low termination rate along the year 2006 to 2013, which in return when a group of people leave the company at once, there is an empty hole in the company to fill up.

Another problem faced by the company is gender imbalance. As mentioned by BrainStation Blog, 34 percent of the participant who worked with men mentioned that their gender is an obstacle for success, as compared to a more balanced environment, 19 percent of the participant mentioned their gender is an obstacle for success. (BrainStationBlog,2019). I believe that one of the main reasons why termination rate is high is due to the rising amount of male employee in the work force. During the year 2006-2014, most of the employees are female but throughout the year, male too is interested in joining the workforce.

Analysis 2: Observation on terminated employees

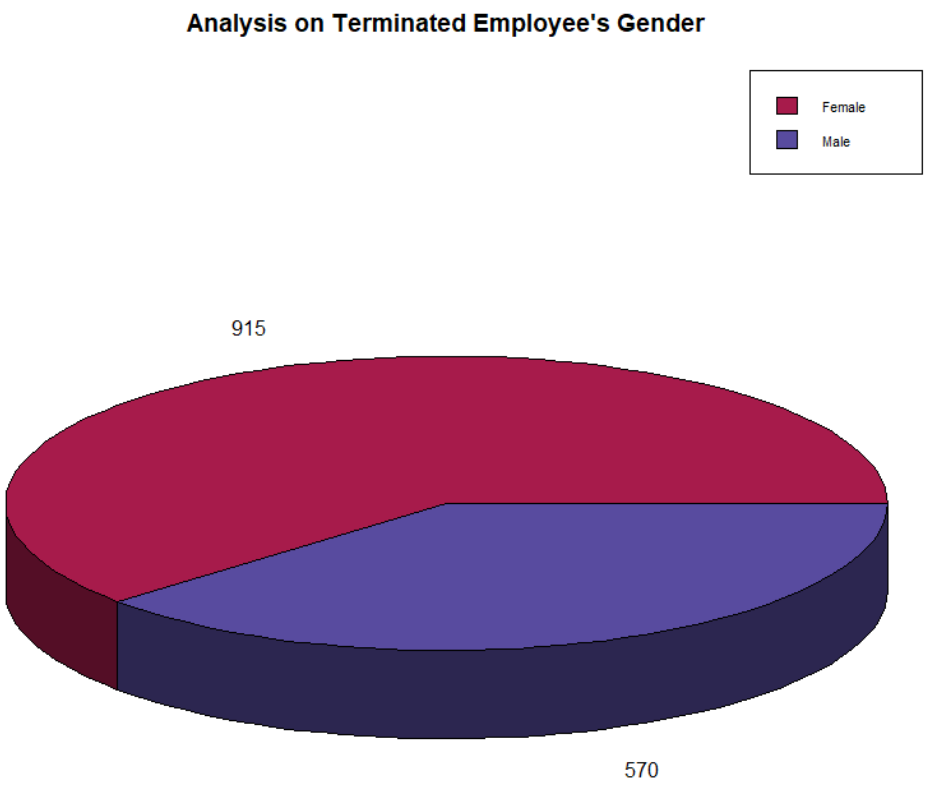
Analysis 2.1: Observation on terminated employees based on their gender

Source code

```
#gender
gen_termi = table(data2 %>% filter(Status =="TERMINATED") %>% select(Gender_Full))
pie3D(gen_termi,col=hcl.colors(length(gen_termi),"Spectral"),
      main="Analysis on Terminated Employee's Gender",border = "black",shade=0.5,labels = gen_termi,labelcex = 1)
legend("topright",c("Female","Male"), cex = 0.7, fill=hcl.colors(length(gen_termi),"Spectral"))
```

line	Explanation
1	The variable gen_termi is used to select and store status which filter terminated employee using the table command while selecting their gender
2	To generate a pie chart
3	To generate legend to state the informaiton of the pie chart

Data Visualization



Observation:

Based on the data visualization above, we could observe that mose of the employee who are terminated are mostly female.

This shows that the working environment does not attract in retaining female employee which cause the amount of terminated female employee be more than male employee. Other reason could varies such as female may choose to resign due to being the caretaker of the family, or problems faced during work.

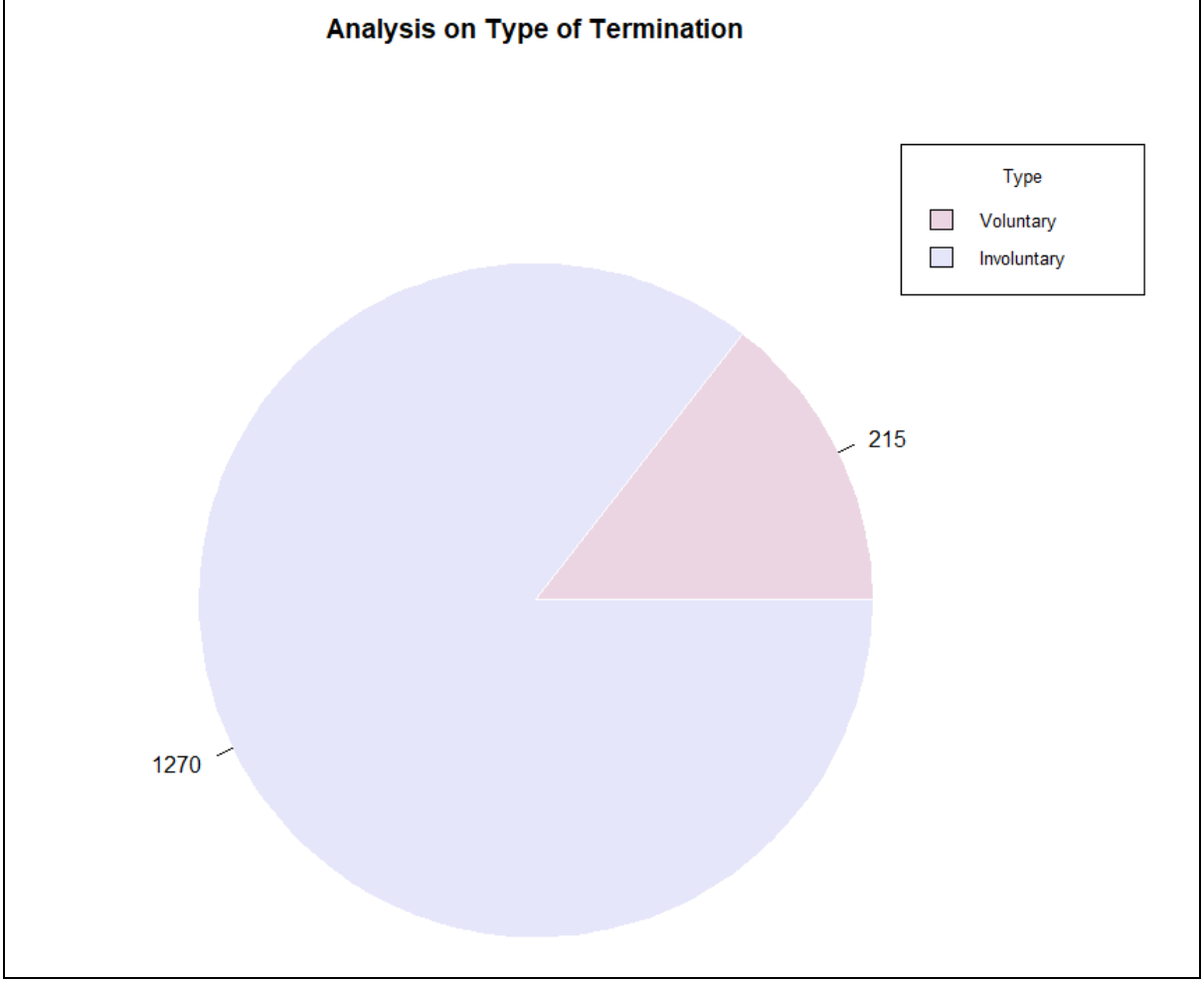
Analysis 2.2: Observation on terminated employees based on termination type

Source code

```
#type
type_termi = table(data2 %>% filter(Status =="TERMINATED") %>% select(Type_of_Termination))
pie(type_termi,label=type_termi,main = "Analysis on Type of Termination",col=c("#ecd5e3","lavender"),border="white" )
legend("topright",c("Voluntary","Involuntary"), cex =0.8, fill=c("#ecd5e3","lavender"),title="Type")
```

line	Explanation
1	The variable type_termi is used to select and store status with the filter terminated employee using the table function while selecting the type of termination.
2	To generate a pie chart
3	To generate legend to state the informaiton of the pie chart

Data Visualization



Observation:

Based on the data visualization above, we could observe that most of the terminated employees are Involuntary, which means that 85% of the terminated employees are laid off, whereas the other 15% of terminated employees are due to resignation or retirement.

This shows that the company's high rate of involuntary termination are sought to be the reason of the company poor hiring and onboarding practices

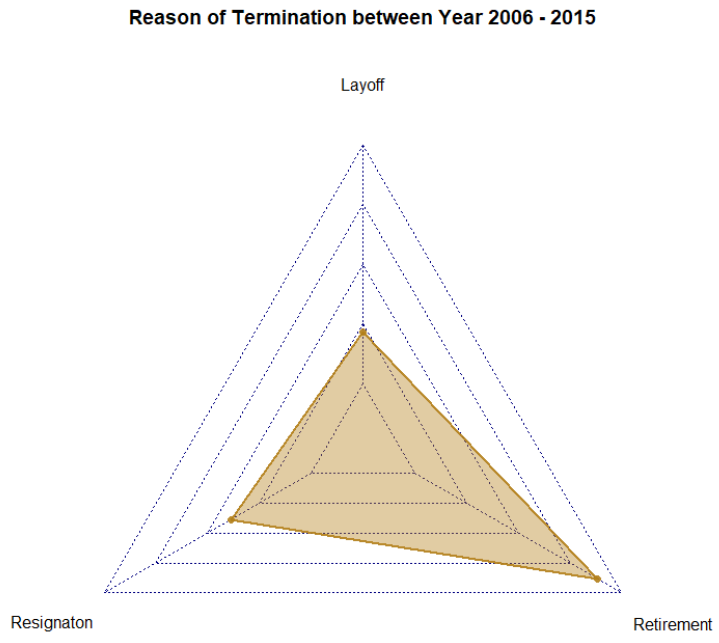
Analysis 2.3: Observation on terminated employees based on their termination reason

Source code

```
#reason
layoff_t = matrix(rbind(max=1000,min=0,values=as.integer(data2 %>% filter( Reason_of_Termination=="Layoff") %>%
                                                         summarise(Layoff=n())))
resign_t = matrix(rbind(max=1000,min=0,values=as.integer(data2 %>% filter( Reason_of_Termination=="Resignaton") %>%
                                                         summarise(Layoff=n()))))
retire_t = matrix(rbind(max=1000,min=0,values=as.integer(data2 %>% filter( Reason_of_Termination=="Retirement") %>%
                                                         summarise(Layoff=n()))))

terminate_emp = data.frame(cbind(layoff_t,resign_t,retire_t))
radarchart(terminate_emp,title = "Reason of Termination between Year 2006 - 2015",
           vlabels = c("Layoff","Resignaton","Retirement"),
           pcol=rgb(0.7,0.5,0.1,0.9) , pfcol= rgb(0.7,0.5,0.1,0.4) ,plwd=2)
```

line	Explanation
1	Assigning the variable layoff_t the matirx of reason of termination whih filter empoyees who are layoff using the rbind function
2	Assigning the variable resign_t the matirx of reason of termination whih filter empoyees who are resign using the rbind function
3	Assigning the variable retire_t the matirx of reason of termination whih filter empoyees who are retire using the rbind function
4-5	Generating the radar chart
5	Creating the title and the label of the radar chart

Data Visualization



Observation:

Based on the data visualization above, we could see that retirement holds the majority of the reasons for why employees leave the company from the year 2006 to year 2015, whereas resignation holds the second majority of the reasons for why employees leave the company.

This shows that most of the company workforce are older employees thus resulting in higher retirement as compared to other reasons like resignation and layoff.

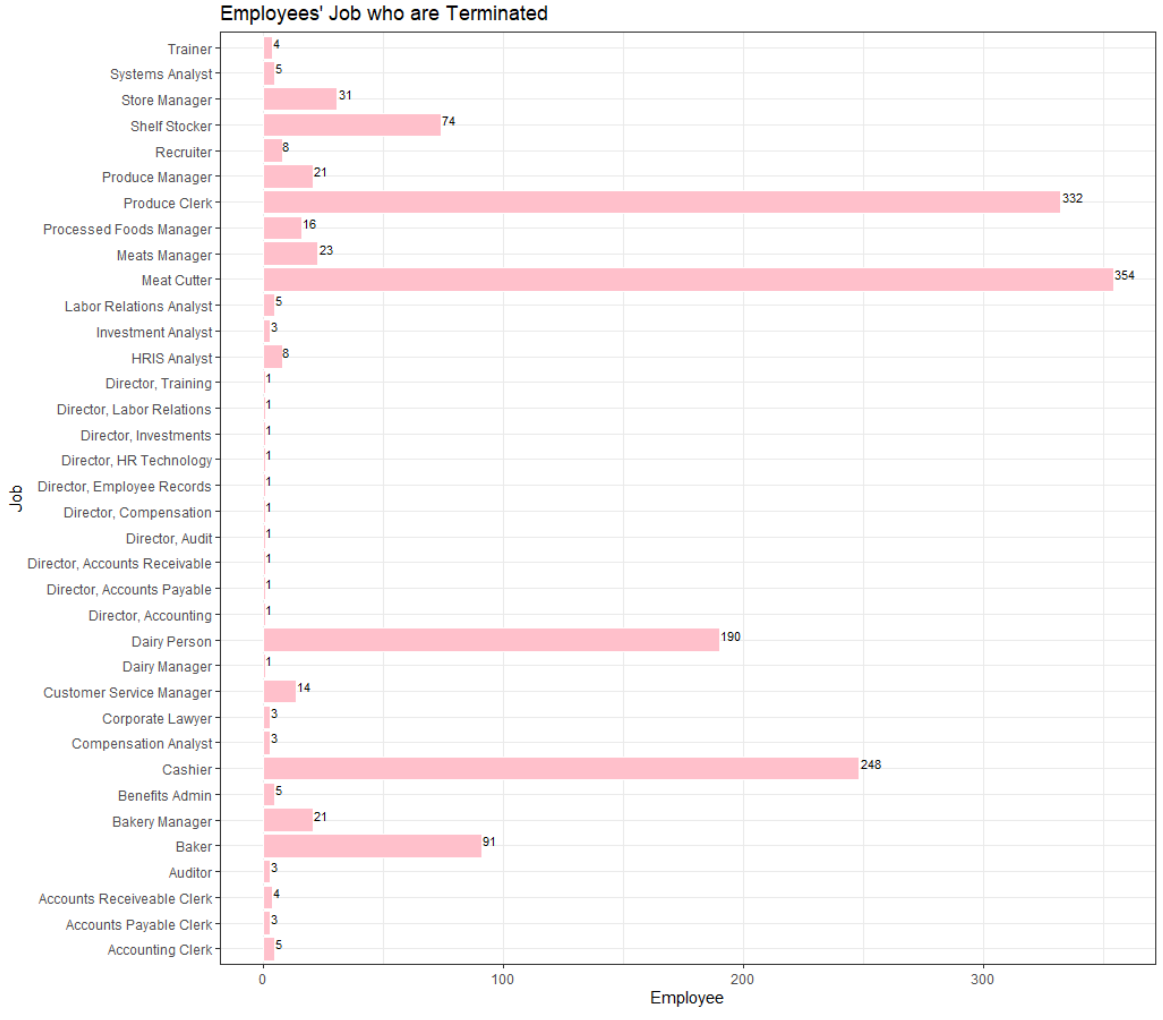
Analysis 2.4: Observation on terminated employees based on their job

Source code

```
#job
data2 %>% filter(Status == "TERMINATED") %>% group_by(Job) %>% summarise(Employee = n()) %>%
  ggplot(aes(x=Job,y = Employee))+geom_bar(stat="identity",fill = "pink",col="white")+
  coord_flip()+ggtitle("Employees' Job who are Terminated")+
  geom_text(aes(label=Employee),hjust=-0.1,vjust=0,col="black",size=3)+theme_bw()
```

line	Explanation
1	Summarize the data by selecting employees with status filter to only select terminated employees which is grouped by their job.
2-4	Generate bar chart
3	Generate title for the bar chart

Data Visualization



Observation:

Based on the data visualization above, we could see that most of the terminated employees used to work for the company as produce clerks and meat cutters. We could also see a moderate amount of terminated employees are dairy person, cashier,baker and shelf stocker.

This shows that most low level skills jobs are not ideal in retaining employees. In my opinion, employees who worked a low level skill job choose to leave the company due to various reasons, one of them is to look for higher salary, whereas second is to search for the needs to be valued.

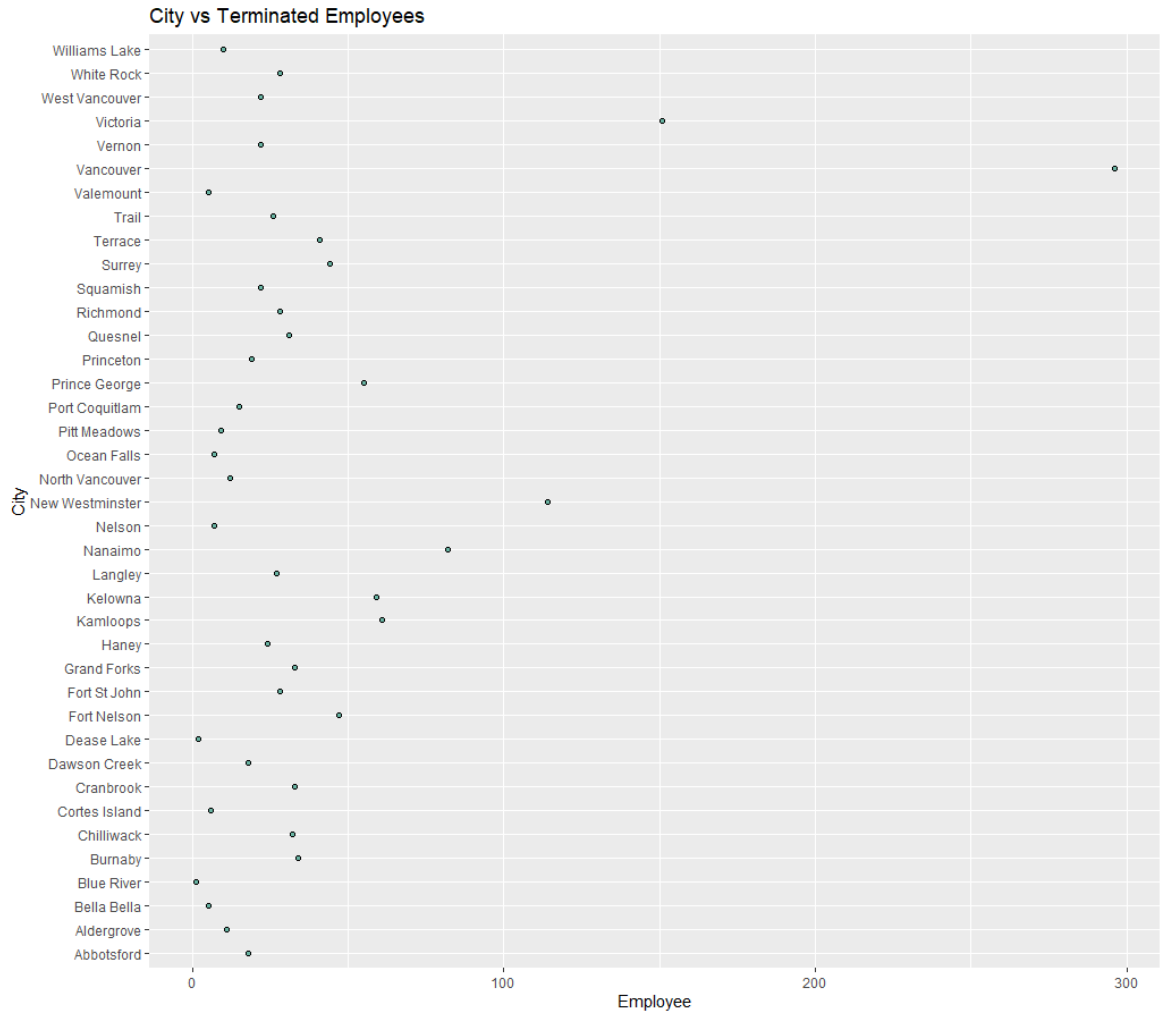
Analysis 2.5: Observation on terminated employees based on their city

Source code

```
#city
city_t = data2 %>% filter(Status == "TERMINATED") %>% group_by(City)%>% summarise(Employee = n())

city_t %>% arrange(Employee) %>% ggplot(aes(x=Employee,y = City))+geom_point(shape=21,col="black",fill="#69b3a2")+
  ggtitle("City vs Terminated Employees")
```

line	Explanation
1	Assigning the variable city_t with summarized data status filter to terminated employee and the group by city
2	Generate scatterplot graph

Data Visualization



Observation:

Based on the data visualization, we could see that Vancouver had the most total terminated employee throughout the year 2006 to the year 2015. After Vancouver, Victoria is the second city in terms of cities from which the terminated employees are from.

This shows that there are more terminated employees from cities like Vancouver, Victoria and New Westminster, due to the population size there. These three cities are considered Canada's major cities, which house most of Canada's population. Thus there are more working opportunities there as compared to other rural areas.

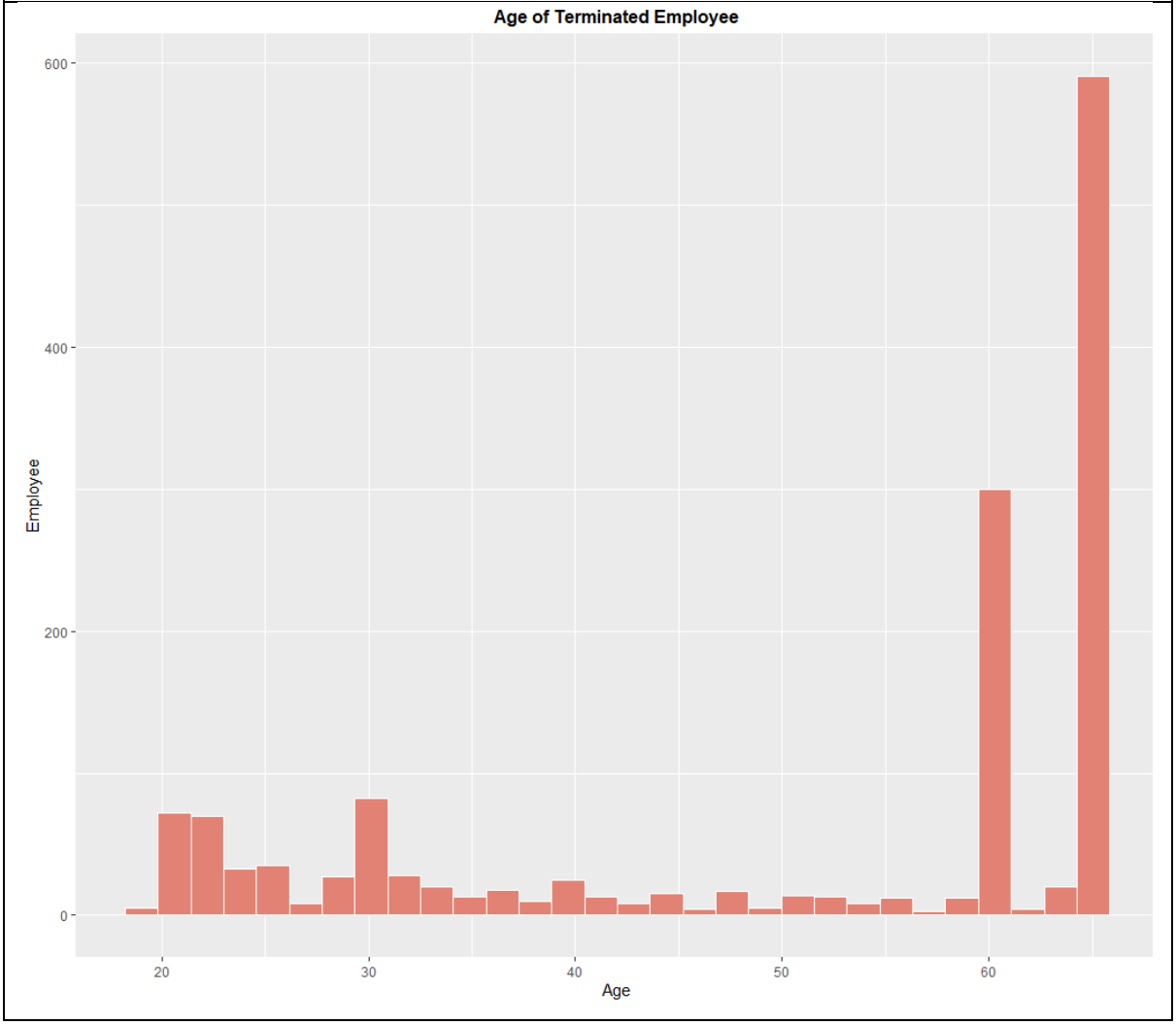
Analysis 2.6: Observation on terminated employees based on their age

Source code

```
#age
data2 %>% filter(Status == "TERMINATED") %>% group_by(Age)  %>% ggplot(aes(x=Age)) +
  geom_histogram(fill="#e28274",col="white")+ggtitle("Age of Terminated Employee")+ylab("Employee")+
  theme(plot.title=element_text(size=12,face="bold",hjust = 0.5))
```

line	Explanation
1	Summarize the data status filter to only terminated employee and group by their age.
2-3	To generate the bar chart.

Data Visualization



Observation:

Based on the data visualization above, we could vie that most of the terminated employee are from the age of 60 and above. But with delegate investigation, we could also view that there is an increase in terminated employees at the age of 20 to 30. The amount of employee terminated at the age of 35 to 55 are very low.

This shows that most of the employees either choose to leave the company at a young age to seek better opportunities outside, or resign at an old age.

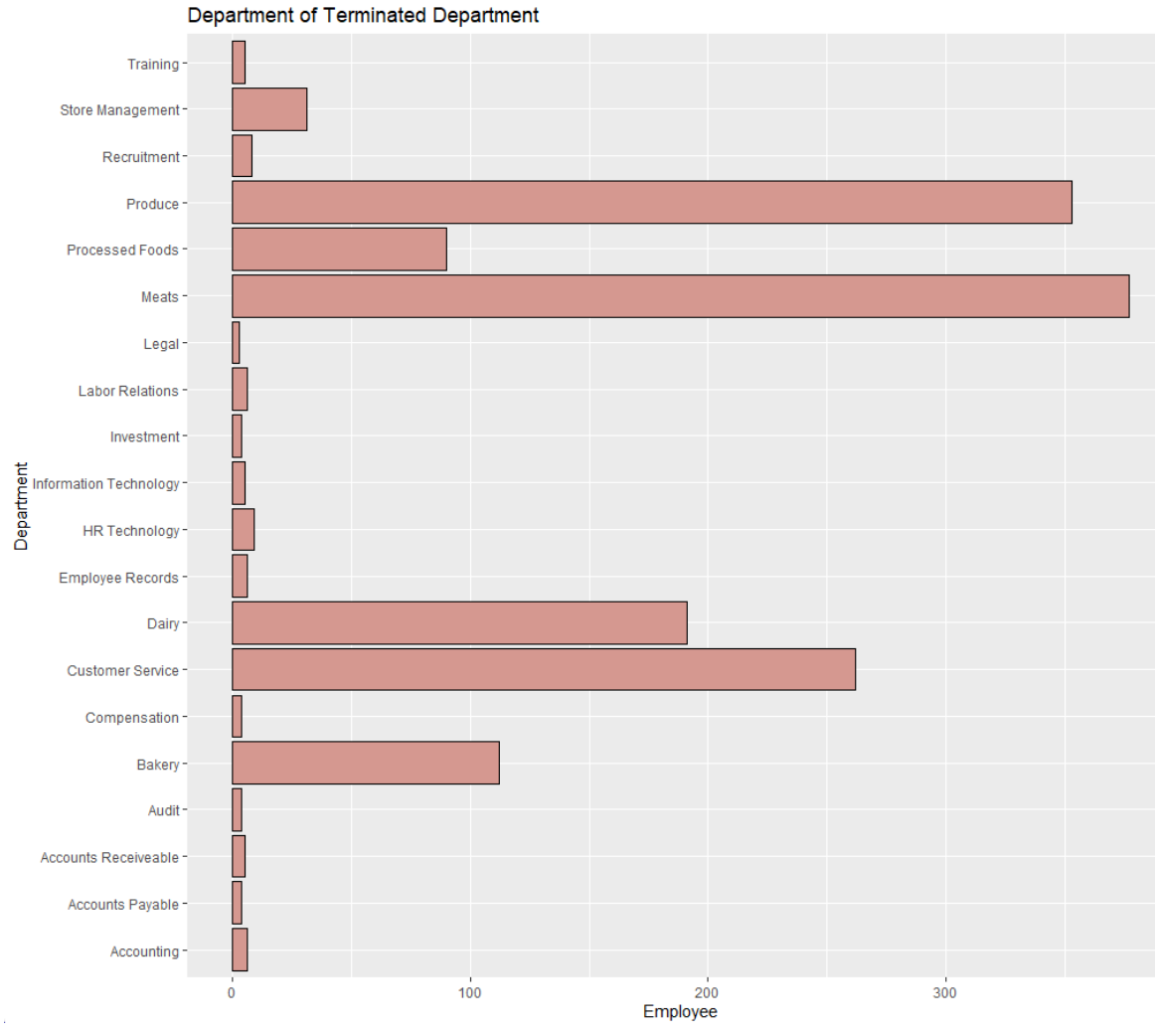
Analysis 2.7: Observation on terminated employees based on their department

Source code

```
#department
data2 %>% filter(Status == "TERMINATED") %>% group_by(Department) %>% summarise(Employee=n()) %>%
  ggplot(aes(x=Department,y=Employee))+geom_bar(stat="identity", position = "dodge",
                                                fill = "#d59890",col = "black")+coord_flip()+
  ggtitle("Department of Terminated Department")
```

line	Explanation
1	Summarize the data status to filter only terminated employee and group by their department
2-3	Generate bar chart
3	Generate title for the bar chart

Data Visualization



Observation:

Based on the data visualization above, we could see that most of the terminated employees are from the produce, meats and customer service department. Other than the top 3 mentioned before, the processed food, dairy and bakery department also shows a high amount of terminated employees.

This shows that low level labor intensive work is not preferable by the employee who worked there, which may result in a high termination rate among those departments.

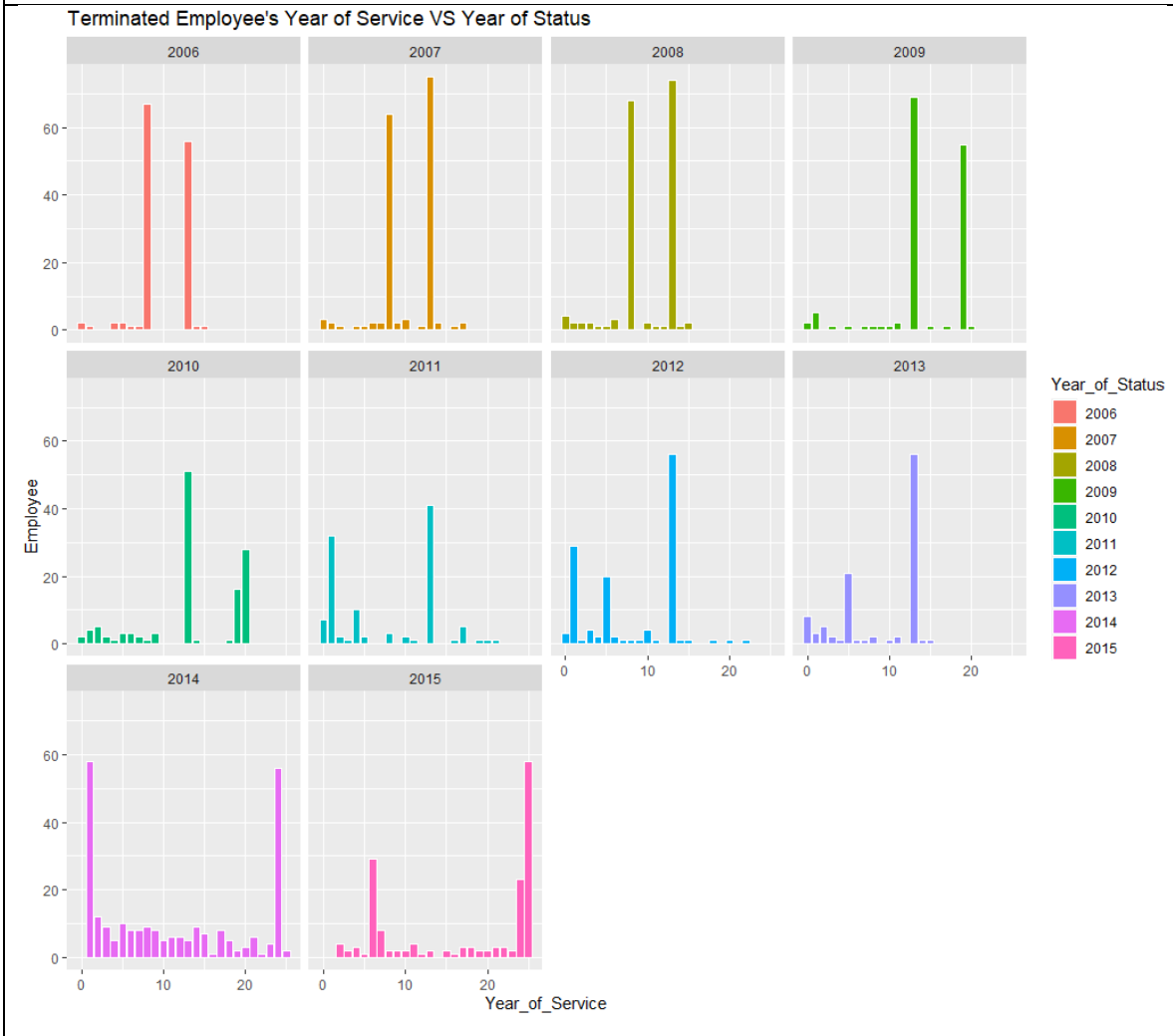
Analysis 2.8: Observation on terminated employees based on their length of service

Source code

```
#length
data2 %>% filter(Status == "TERMINATED") %>% group_by(Year_of_Service,Year_of_Status) %>% summarise(Employee=n()) %>%
  ggplot(aes(x=Year_of_Service,y=Employee,fill=Year_of_Status))+geom_bar(stat="identity",col="white")+
  facet_wrap(~Year_of_Status)+
  ggtitle("Terminated Employee's Year of Service VS Year of Status")
```

line	Explanation
1	Summarize the data status to filter terminated employee then group by the year of service and year of status.
2	To generate bar chart
3	To create multiple panel
4	To generate the title for the bar chart

Data Visualization



Observation:

Based on the data visualization above, we could see an obvious trend happening throughout the year 2006 to year 2015.

From 2006 to 2008, we could observe that most terminated employees are from the middle age group, probably from the age 25 to 40 years old.

From 2009 to 2010, we could observe that there is an increase of termination of senior employees in the company.

From 2011 to 2013, the termination of employees from the middle age group maintained but the senior age group decreased, however there is an increase of termination of the younger group of employees.

From 2014 to 2015, the termination of employees from the middle age group has stopped but the senior age group has seen an increase and the termination of the younger group of employees has also increased.

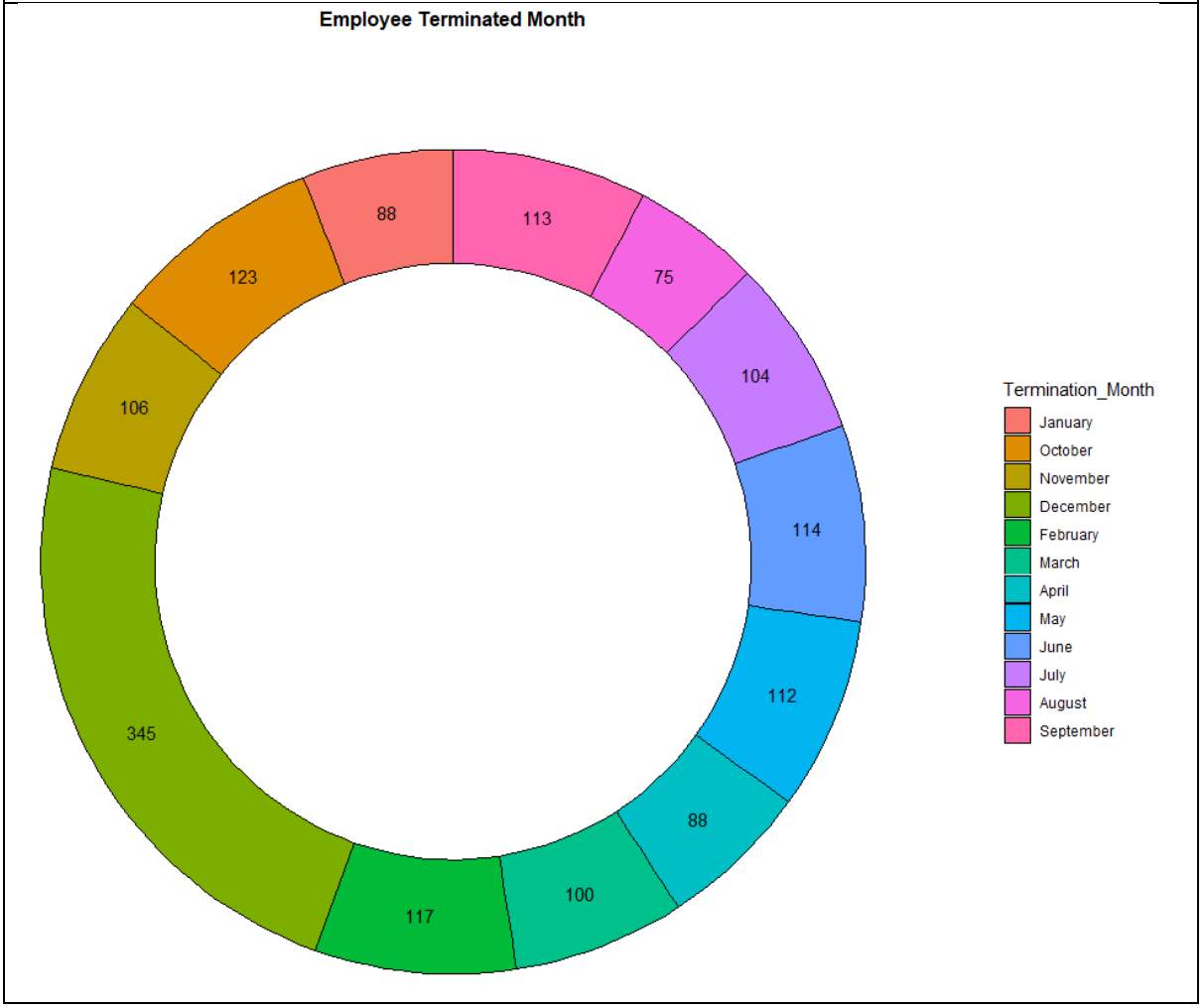
Analysis 2.9: Observation on terminated employees based on the month of the year

Source code

```
#month
data2 %>% filter(Status == "TERMINATED") %>% group_by(Termination_Month) %>% summarise(Employee=n()) %>%
  ggplot(aes(x=3,y=Employee,fill=Termination_Month))+geom_col(col="black")+coord_polar(theta="y")+
  xlim(c(0.2,3.5))+theme(panel.background = element_rect(fill = "white"),
                         axis.title = element_blank(),axis.ticks = element_blank(),axis.text = element_blank())+
  geom_text(aes(label=Employee),position = position_stack(vjust =0.5))+ ggtitle("Employee Terminated Month")+
  scale_fill_discrete(labels=c("January","October","November","December","February","March","April","May",
                               "June","July","August","September"))+
  theme(plot.title=element_text(size=12,face="bold",hjust = 0.5))
```

line	Explanation
1	Summarize the data status which filters terminated employees and group by their termination month
2-6	Generate the donut chart
4	Generate lagend to shwcase the info of the donut chart

Data Visualization



Observation:

Based on the data visualization above, we could see that most of the employees leave the company at december. As most of the terminated employees are terminated at the month of december. Except for December, the ratio of terminated employees for the rest of the month are almost equal to each other.

This shows that the company usually does their employee performance review at the month of December which results in a lot of employees being terminated at that certain month.

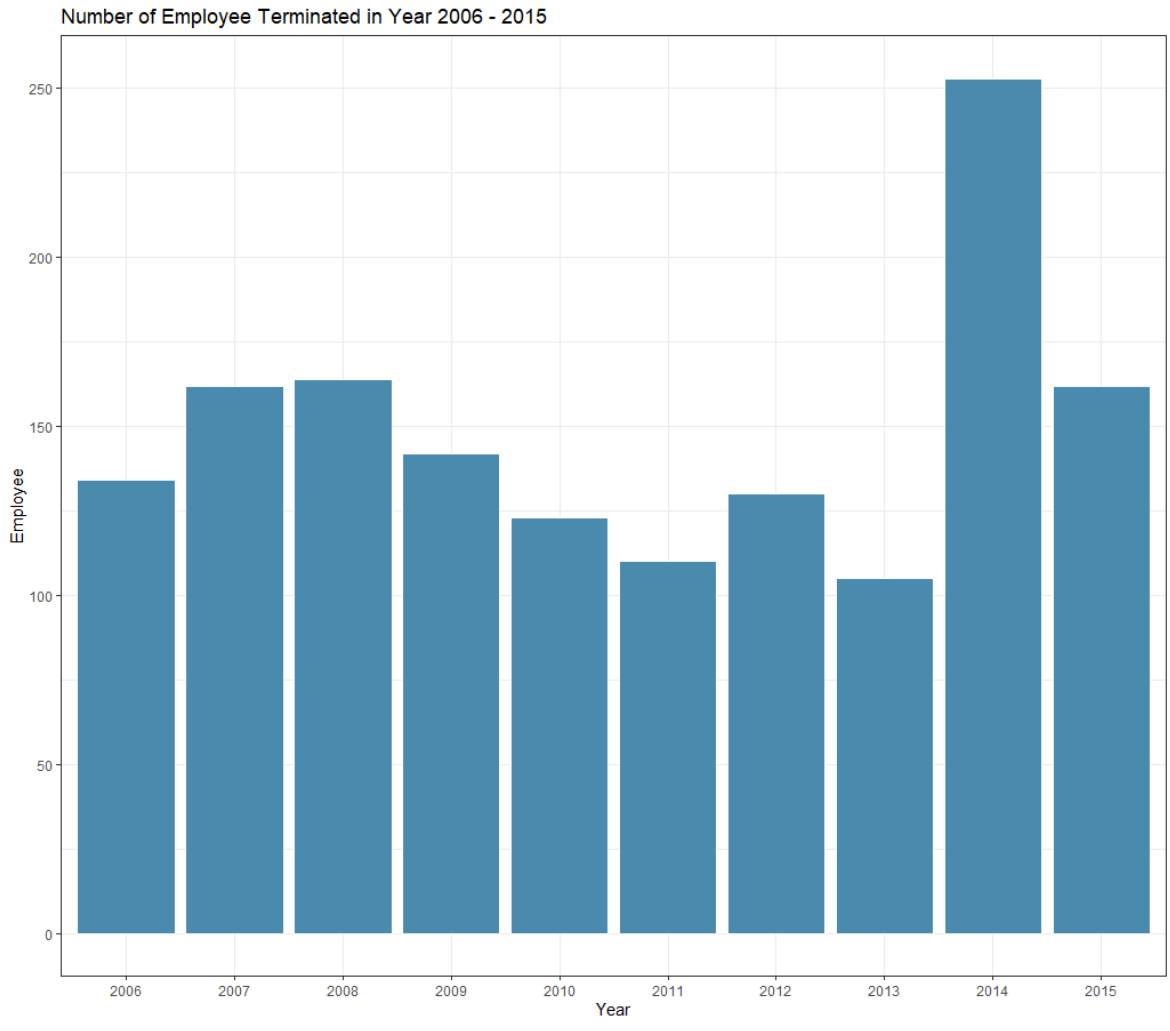
Analysis 2.10: Observation on terminated employees based on each year

Source code

```
#year of status
data2 %>% filter(Status == "TERMINATED") %>% group_by(Year_of_Status) %>% summarise(Employee=n()) %>%
  ggplot(aes(x=Year_of_Status,y=Employee))+geom_bar(stat="identity",fill = "#4a8bad",col = "white")+theme_bw()+
  ggtitle("Number of Employee Terminated in Year 2006 - 2015")+xlab("Year")
```

line	Explanation
1	Summarize the data status to filter terminated employee then group them by year of status
2	Generate bar chart
3	Generate title for the bar chart

Data Visualization



Observation:

Based on the data visualization above, we could see that in the year 2014, there is a spike in employee termination, which indicates something has happened in the company. There is a decrease in termination before 2014, and the years 2007,2008 and 2015 have moderate termination.

This shows that some event may have happened in the year 2014 which results in a spike in terminated employees. Other suggestions are that the spike in 2014 may be the cause of the low termination rate during the year 2009-2013.

Conclusion of Analysis 2: Observation on terminated employees

After conducting various analysis based on: -

- Observation on terminated employees based on their gender
- Observation on terminated employees based on termination type
- Observation on terminated employees based on their termination reason
- Observation on terminated employees based on their job
- Observation on terminated employees based on their city
- Observation on terminated employees based on their age
- Observation on terminated employees based on their department
- Observation on terminated employees based on their length of service
- Observation on terminated employees based on the month of the year
- Observation on terminated employees based on each year

It could be concluded that: -

- Retirement holds the majority of the reason of why employee leave the company
- Cities like Vancouver, Victoria and New Westminster have high amount of terminated employees.
- Most terminated employees are terminated involuntary
- Most terminated employees are between the age 20-30 and 60 and above
- Most of the terminated employees are from the produce, meats, customer service , processsed food, dairy and bakery department
- From year 2006-2008, we could observe that most terminated employees are from the middle age group, probably from the age 25 to 40 years old.
- From year 2009-2010, we could observe that there is an increase of termination of senior employees in the company.
- From year 2011-2013, the termination of employees from the middle age group maintain but the senior age group have decrease however there is an incease of termination of the younger group of employees.
- From year 2014 to 2015, the termination of employees from the middle age group have stop but the senior age group have seen an increase and the the termination of the younger group of employees are also increased
- The month December has the highest termination rate

- The termination rate decreases along the year 2009-2013, but have a steep increase on year 2014

Evidence: -

It is mentioned by indeed.com, where some companies conduct evaluations when the company reaches the end of the year, employee who excels at the evaluation are typically removed from probationary employment issues (Indeed.com,2021). Relating back to the company, the company is suspected to conduct their evaluation in November and December causing high termination rate. Other than that, most of the reason of terminated employee are involuntary, which suspects that the evaluation plays a huge role in employees' termination.

It is mentioned by nestegg.com, where 43% of men and 66% of women retired from the workforce due to unable to achieve daily quota regardless of their intention to stay (Nestegg, 2019). Relating back to the company, suspect that senior employees are being forced to retire due to decline in productivity, moreover, factors like health and family too plays a role.

Analysis 3: Observation on the gender ratio in the company since year 2006 to year 2015

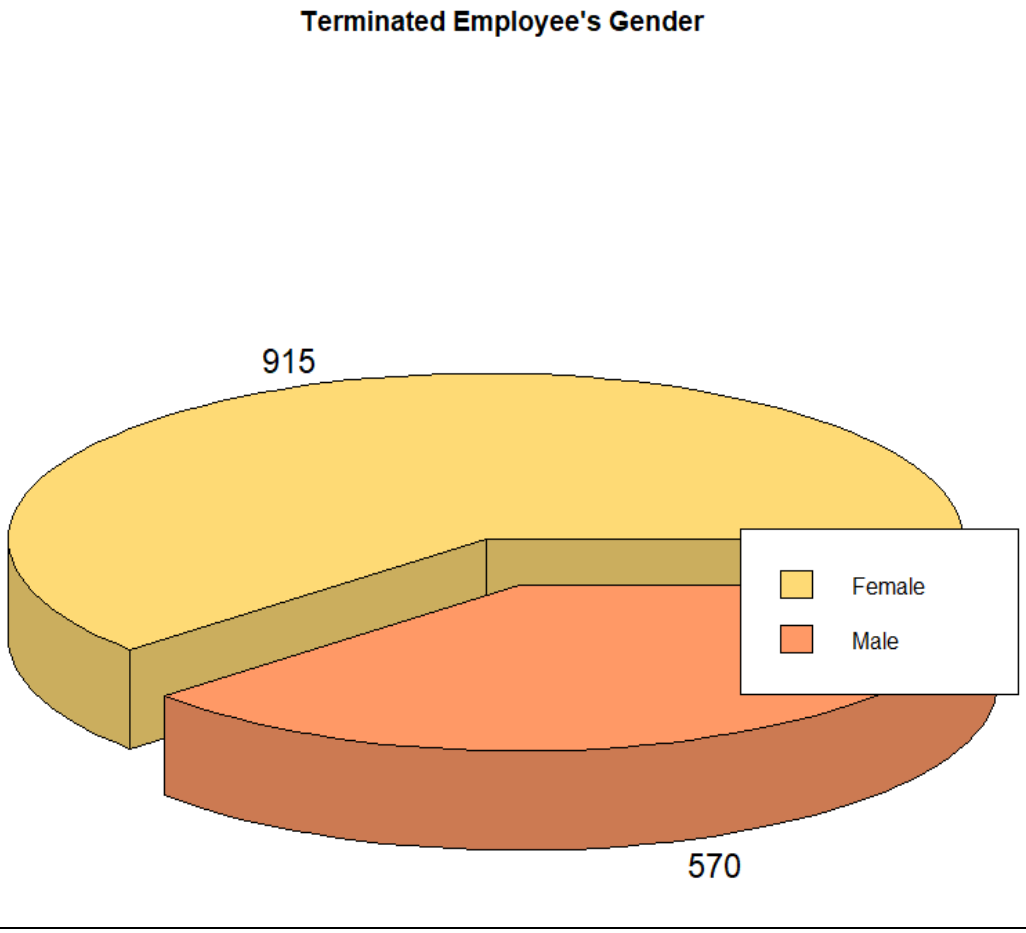
Analysis 3.1: Observation on the gender ratio of the company's terminated employees

Source code

```
#gender
gender_t = table(data2 %>% filter(Status == "TERMINATED") %>% select(Gender_Full))
pie3D(gender_t,labels=gender_t,cex=0.7,radius=1,main="Terminated Employee's Gender",border="black",
      col=c("#feda75","#ff9966"),explode = 0.1)
legend("right",c("Female","Male"),cex=1,fill=c("#feda75","#ff9966"))
```

line	Explanation
1	The variable gender_t is assign to select and store status with the filter of terminated employee using the table function while selecting their gender.
2	To generate a 3D pie chart
3	To generate legend to show information of the pie chart

Data Visualization



Observation:

Based on the data visualization above, we could observe that most of the employees who are terminated are mostly female.

This shows that the working environment does not attract female employees which causes the amount of terminated female employees to be more than male employees.

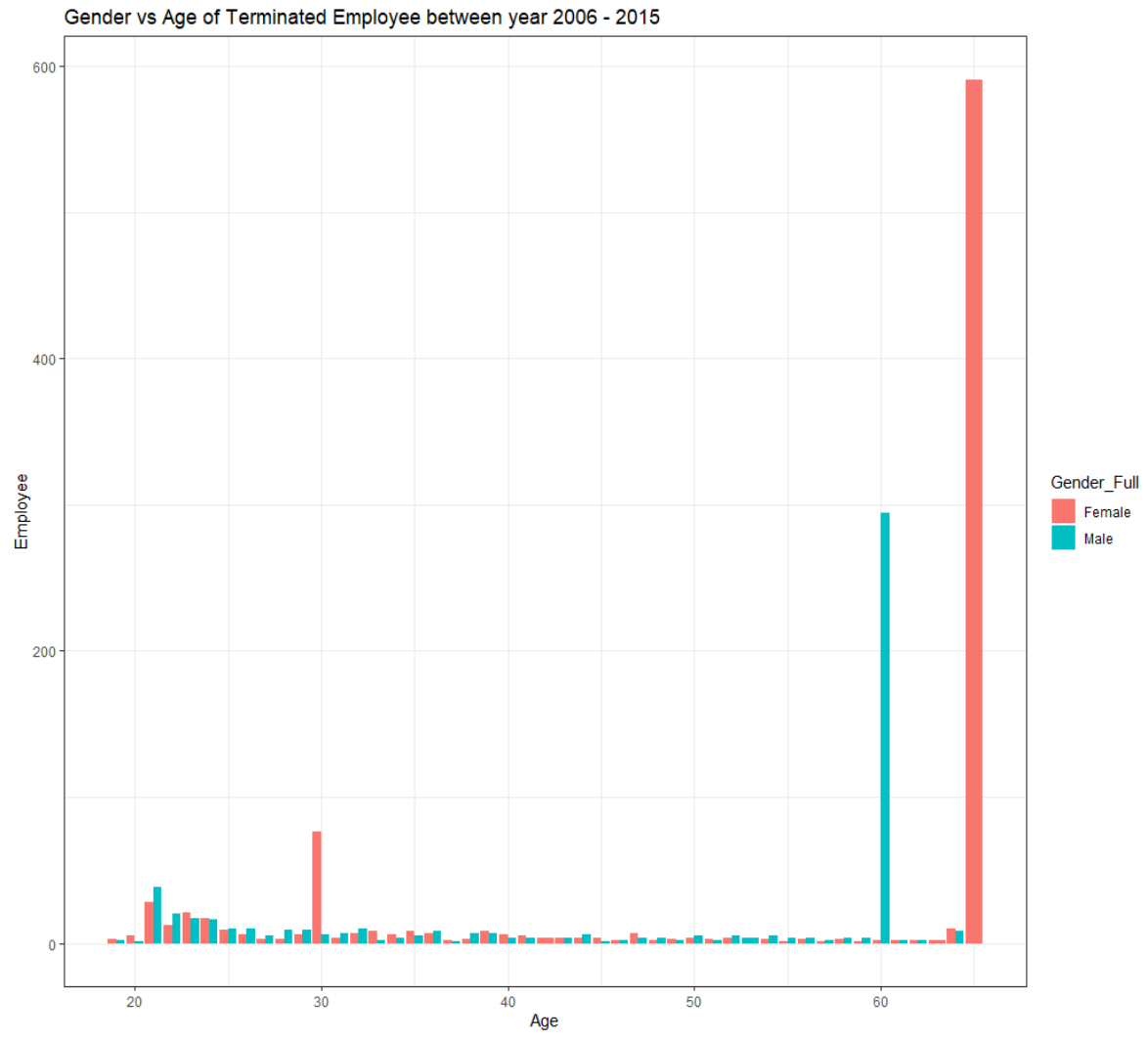
Analysis 3.2: Observation on the gender ratio of the company's terminated employees based on their gender and age

Source code

```
#gender + age
data2 %>% filter(Status == "TERMINATED") %>% group_by(Gender_Full,Age) %>% summarise(Employee = n()) %>%
  ggplot(aes(x=Age,y=Employee,fill=Gender_Full))+geom_bar(stat="identity",position="dodge")+
  ggtitle("Gender vs Age of Terminated Employee between year 2006 - 2015")+theme_bw()
```

line	Explanation
	TO summarize the data status to fitler only tetrinated employee while group them by their gender and age
2	To generate the bar chart
3	To generate title for the bar chart

Data Visualization



Observation:

Based on the data visualization above, we could view that most of the terminated employees are from the age of 60 and above. We also found out that female employees tend to retire later than male employees. But with delegate investigation, we could also see that there is an increase in terminated employees at the age of 20 to 30, we could see that there is a spike of male employees leaving at 20 but the trend soon decreases. At the age group 30, a lot of female employees are terminated. The amount of employee terminated at the age of 35 to 55 are very low.

This shows that there is a trend of young male employees tend to leave the company at the age of 20 whereas young female employees tend to leave the company at the age of 30. Other than that, it is shown that senior male employees tend to retire earlier than female employees.

Analysis 3.3: Observation on the gender ratio of the company's terminated employees based on their gender, job and age

Source code

```
#gender +job + age
data2 %>% filter(Status =="TERMINATED") %>% group_by(Gender_Full,Job,Age) %>% summarise(Employee =n()) %>%
  ggplot(aes(x=Age,y=Job,col=Gender_Full,size=Employee))+geom_point(shape = 19,alpha=0.5,stroke=1)+
  ggtitle("Terminated Employee's Gender Comparing the Job and Age")
```

line	Explanation
1	To summarize the data status to only filter terminated employee while group them by their gender, job and age
2	To generatte the scatter plot graph
3	To generate the title for the scatter plot

Data Visualization

Observation:

Based on the data visualization above, we could observe that there is a trend where female employees are being terminated at the age of 65, whereas male employees are being terminated at the age of 60. We could also view that throughout the year, male cashier, dairy person and baker are constantly being terminated. Most of the young employees who worked as shelf stocker also faced termination, whereas middle aged meat cutters are constantly facing termination.

This shows that the company men retire earlier than women. Moreover, we could see that jobs like cashier, dairy person, shelf stocker and baker have a hard time retaining employees. Whereas jobs like produce clerk and meat cutter relies on workers productivity, hence lots of middle age employees are mostly terminated, hints that the company requires faster workers to fill the roles.

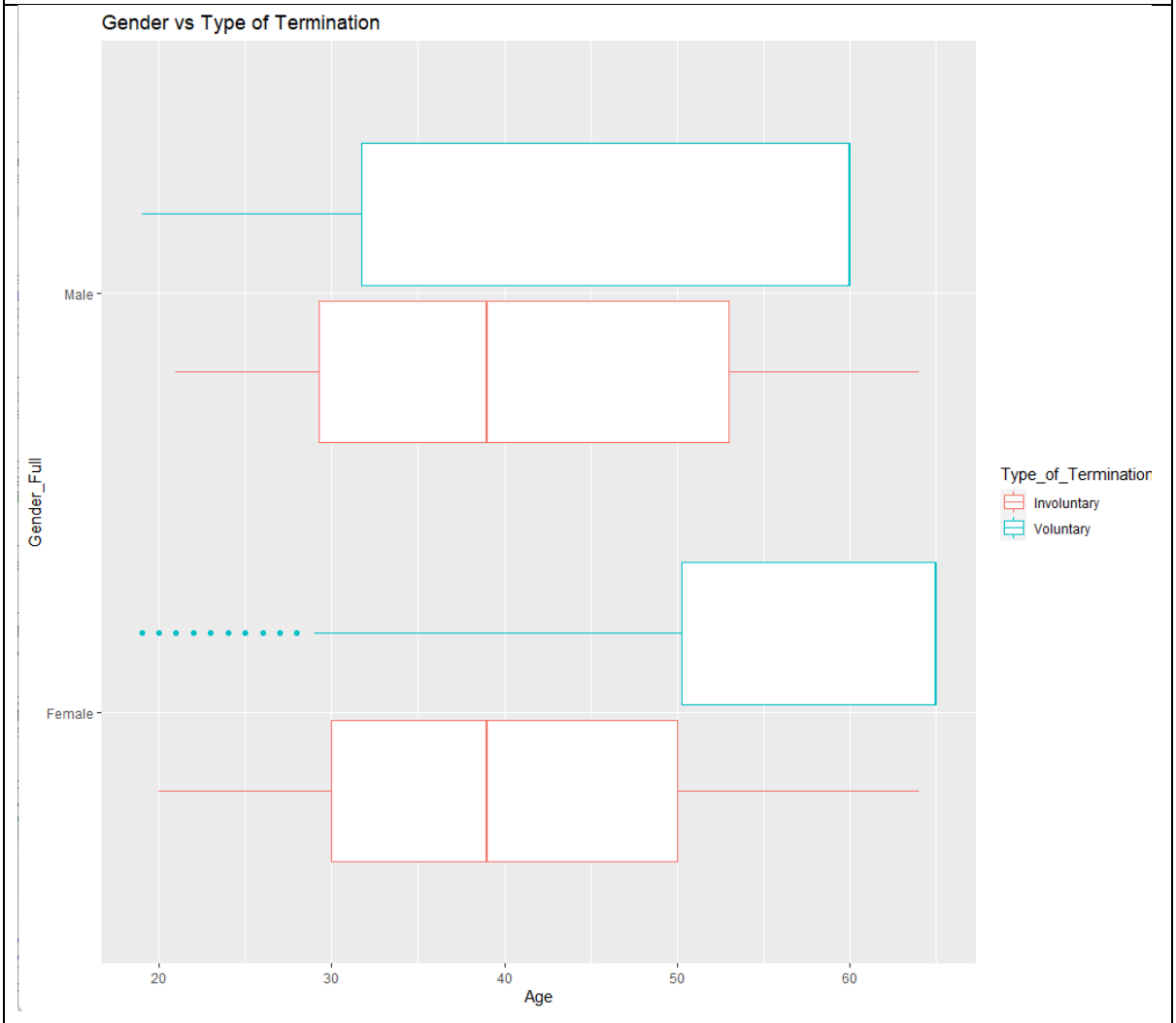
Analysis 3.4: Observation on the gender ratio of the company's terminated employees based on their gender and type of termination

Source code

```
#type + gender
data2 %>% filter(Status == "TERMINATED") %>% group_by(Type_of_Termination,Age,Gender_Full)  %>%
  ggplot(aes(x=Age,y=Gender_Full,col=Type_of_Termination))+geom_boxplot() + ggtitle("Gender vs Type of Termination")
```

line	Explanation
1	To summarize the data status to filter only terminated employee while group them by the type of termination, age and gender.
2	To generate box plot graph

Data Visualization



Observation:

Based on the data visualization above, we could observe that the interquartile of voluntary termination of male employees are between 32 to 60 years old, whereas the interquartile of involuntary termination of male employees are between 30 to 52 years old with a median of 37. The interquartile of voluntary termination of female employees are between 50 to 60 years old, whereas the interquartile of involuntary termination of male employees are between 30 to 50 years old with a median of 37

This shows that most of the terminated male employees who voluntarily terminated are from the age range of 32 to 60 years old, whereas the terminated female employees who voluntarily terminated are from the age range of 50 to 60 years old. The terminated male employees who are involuntarily terminated are from the age range of 30 to 52 years old, whereas the terminated female employees who are involuntarily terminated are from the age range of 30 to 50 years old.

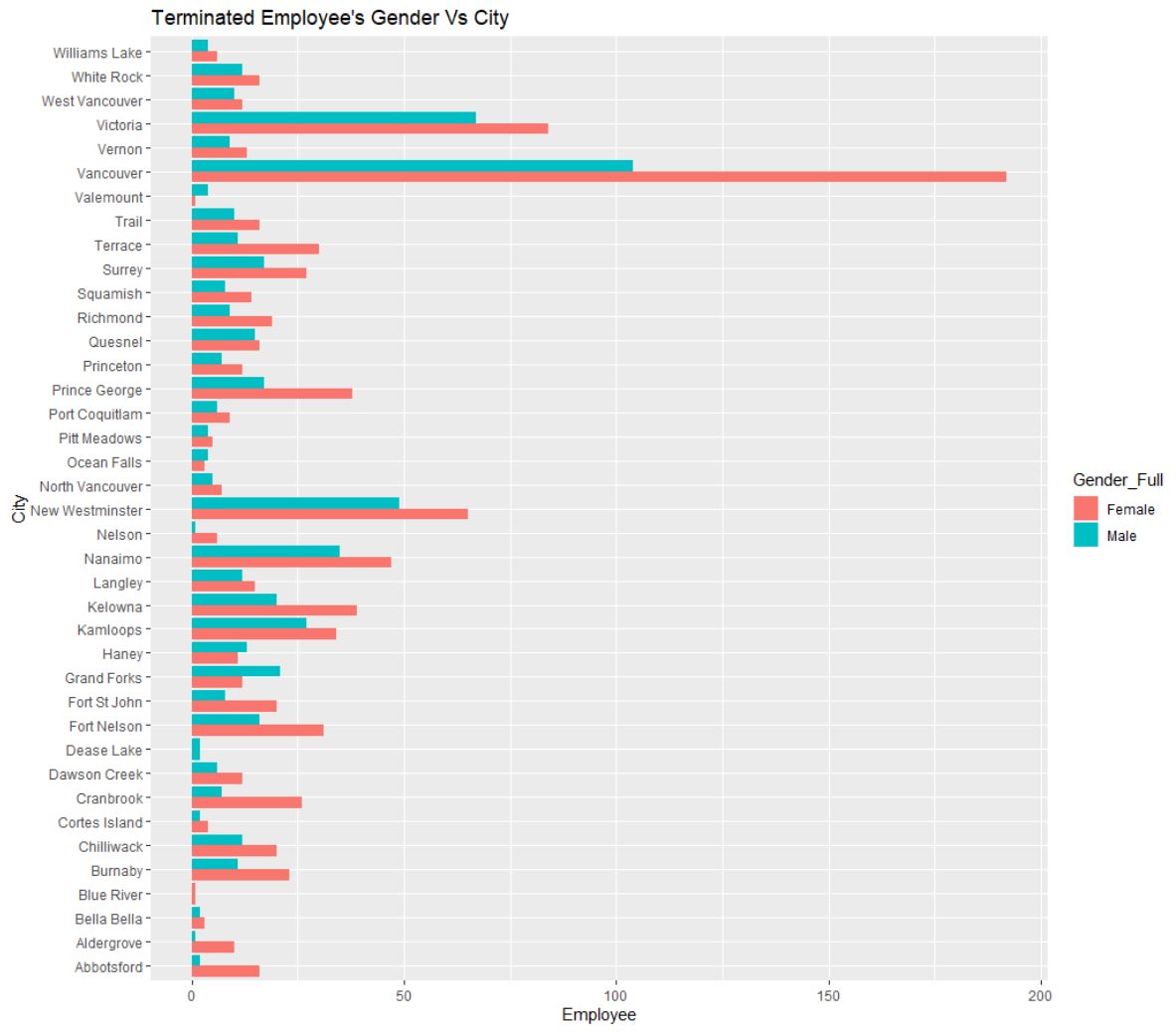
Analysis 3.5: Observation on the gender ratio of the company's terminated employees based on their gender and city

Source code

```
#gender+city
data2 %>% filter(Status == "TERMINATED") %>% group_by(Gender_Full,City) %>% summarise(Employee = n()) %>%
  ggplot(aes(x=City,y=Employee,fill=Gender_Full))+geom_bar(stat="identity",position="dodge")+coord_flip()+
  ggtitle("Terminated Employee's Gender Vs City")
```

line	Explanation
1	To summarize the data status to filter only terminated empployee while group them by their gender and city.
2	To generate the bar chart
3	To generate the title for the bar chart

Data Visualization



Observation:

Based on the data visualization above, we could observe that the majority of the terminated employees are females. By observing the data above, we could see that the amount of female terminated employees at Vancouver is double the amount of the number of male employees terminated.

This shows that most of the termination is conducted at Vancouver and major cities like victoria, new Westminster and Nanaimo.

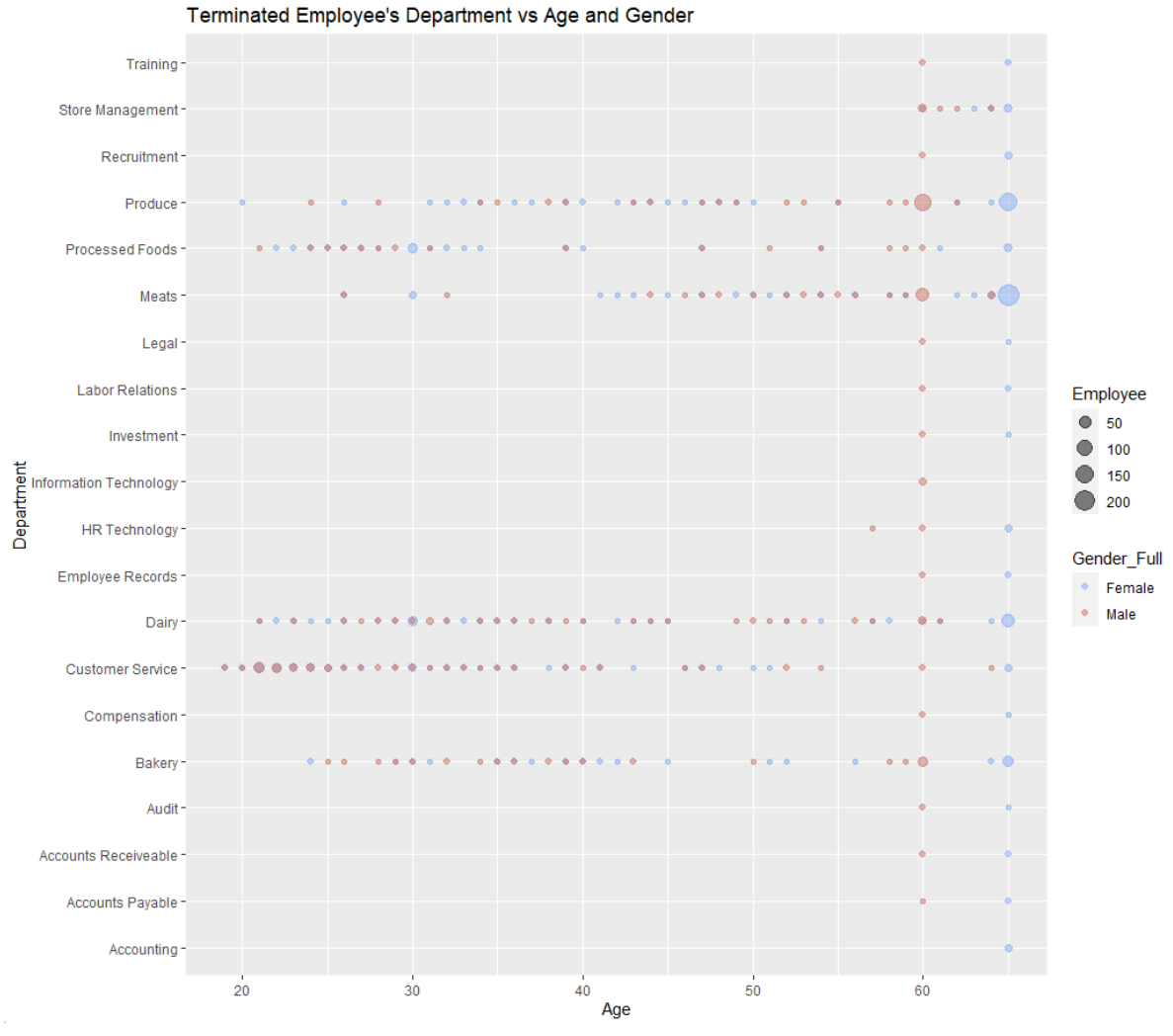
Analysis 3.6: Observation on the gender ratio of the company's terminated employees based on their gender and department

Source code

```
#department + gender
data2 %>% filter(Status =="TERMINATED") %>% group_by(Gender_Full,Department,Age) %>% summarise(Employee =n()) %>%
  ggplot(aes(x=Age,y=Department,col=Gender_Full,size=Employee))+geom_point(shape = 19,alpha=0.5,stroke=1)+
  scale_color_manual(values=c("#84a9f3","#ca7166"))+ ggtitle("Terminated Employee's Department vs Age and Gender")
```

line	Explanation
1	To summarize the data status to filter only terminated empoyee while group them by their gender, departmennt and age.
2-3	To generate the scatter plot graph

Data Visualization



Observation:

Based on the data visualization above, we could observe that there is a trend where female employees are being terminated at the age of 65, whereas male employees are being terminated at the age of 60. We could also view that throughout the year, male customer service, bakery and dairy departments are constantly being terminated. Most of the young employees who worked in the produce and processed foods department also faced termination, whereas middle aged meat departments are constantly facing termination.

This shows that the company men retire earlier than women. Moreover, we could see that departments like customer service, bakery and dairy produce have a hard time retaining employees. Whereas department like produce and meat departments relies on workers productivity, hence lots of middle age employees are mostly terminated hints that the company require faster worker to fill the roles

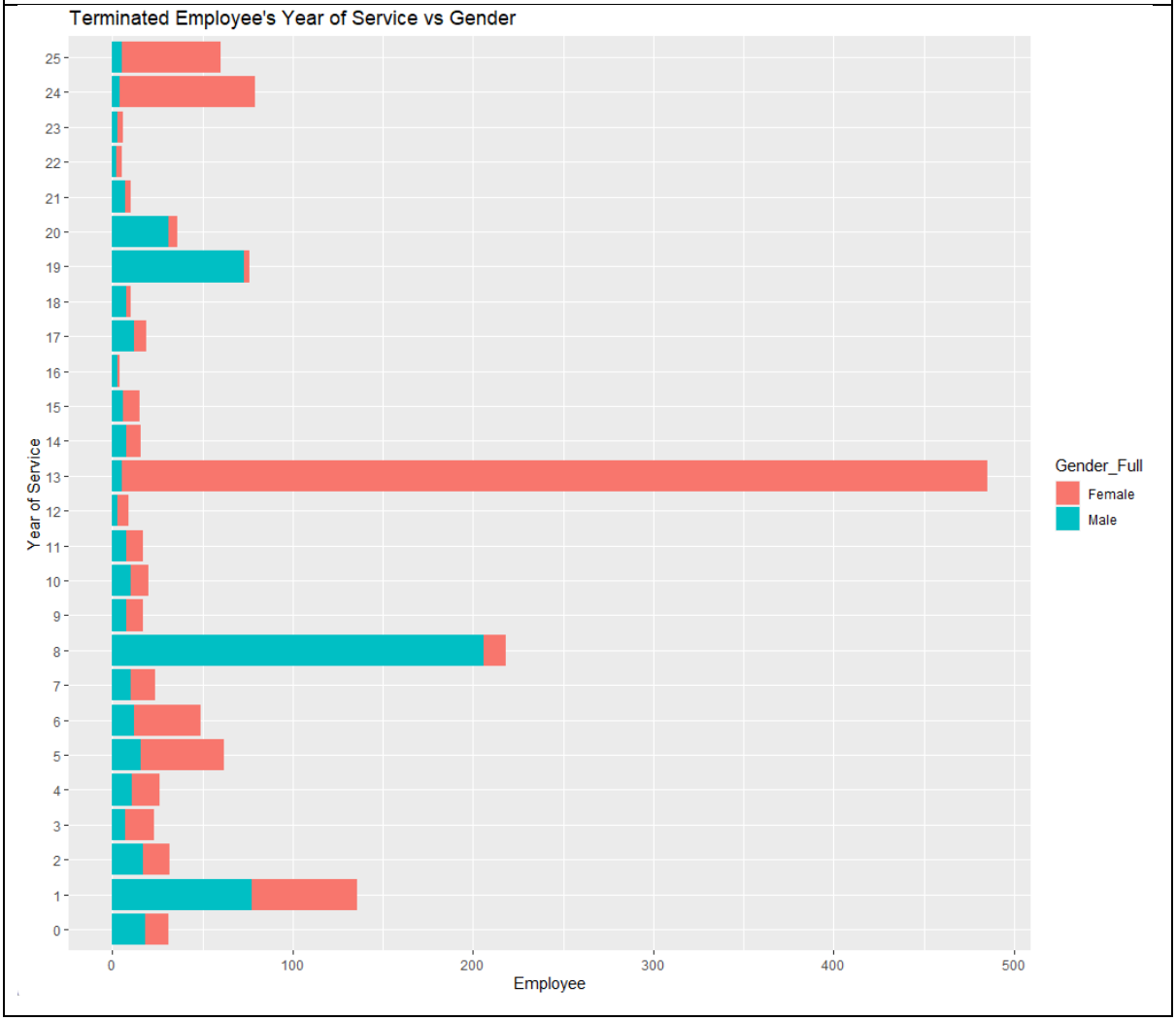
Analysis 3.7: Observation on the gender ratio of the company's terminated employees based on their gender and length of service

Source code

```
#length + gender
data2 %>% filter(Status =="TERMINATED") %>% group_by(Gender_Full,Year_of_Service) %>% summarise(Employee =n()) %>%
  ggplot(aes(x=factor(Year_of_Service),y=Employee,fill=Gender_Full)) + geom_bar(stat="identity")+xlab("Year of Service")+
  coord_flip()+ggtitle("Terminated Employee's Year of Service vs Gender")
```

line	Explanation
1	To summarize the data status to filter only terminated empoyee while grouping them by their gender and year of service
2	To generate the bar chart
3	To generate title for the bar chart

Data Visualization



Observation:

Based on the data visualization above, we could observe that most female workers are terminated after 13 years of working whereas most male workers are terminated after 8 years of service. Moreover, we could see that an amount of male workers are terminated after 19 to 20 years of working, whereas an amount of female workers are terminated after 24 to 25 years of working.

This shows that females tend to serve a longer service year compared to men. Male retire earlier than men, male would choose to leave the company earlier compared to female.

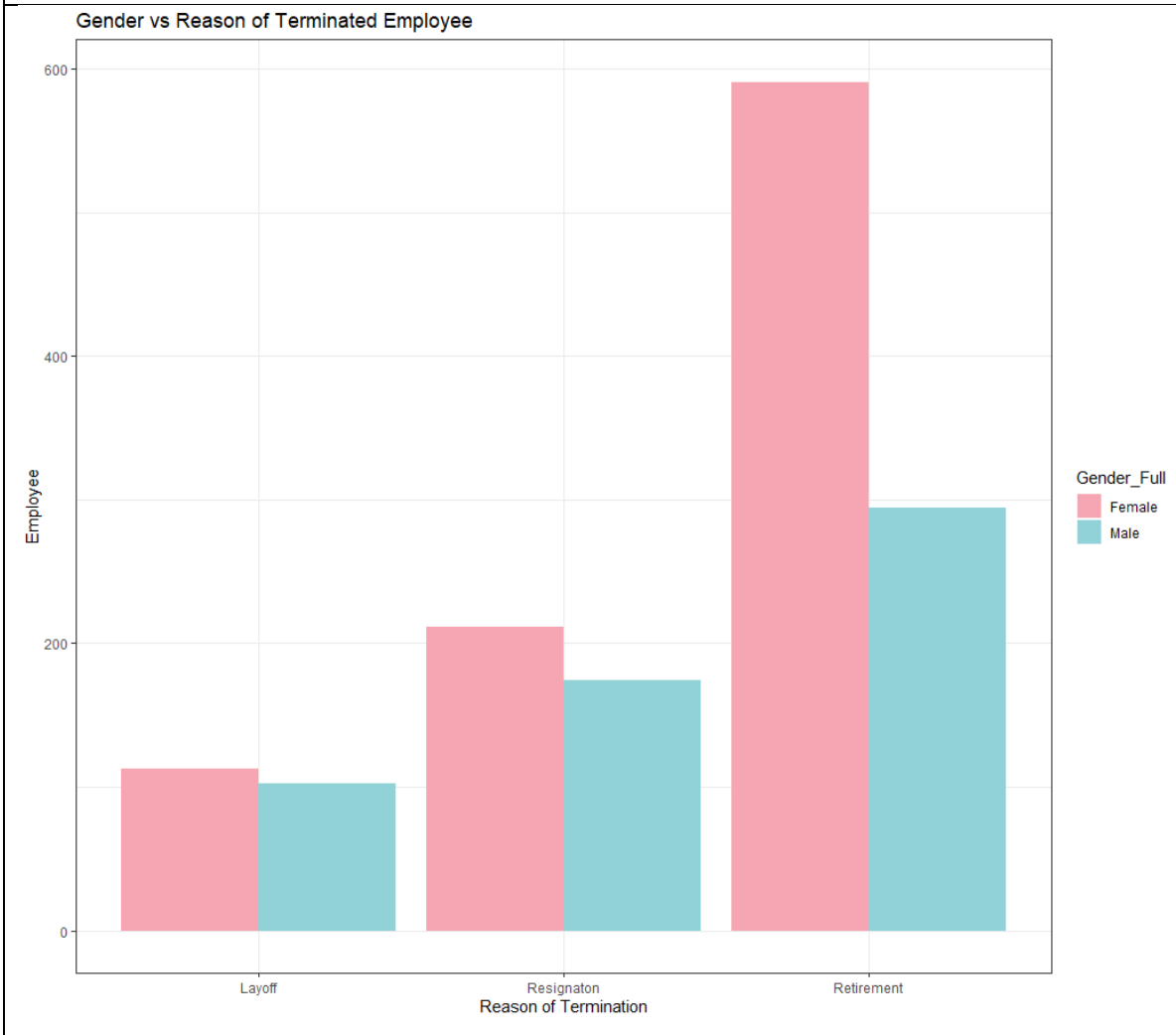
Analysis 3.8: Observation on the gender ratio of the company's terminated employees based on their gender and reason

Source code

```
#gender + reason
data2 %>% filter(Status == "TERMINATED") %>% group_by(Gender_Full,Reason_of_Termination) %>% summarise(Employee = n()) %>%
  ggplot(aes(Reason_of_Termination,y=Employee,fill=Gender_Full))+geom_bar(stat="identity",position="dodge")+
  ggtitle("Gender vs Reason of Terminated Employee")+theme_bw()+scale_fill_manual(values=c("#f6a6b2","#90d2d8"))+
  xlab("Reason of Termination")
```

line	Explanation
1	To summarize he data status to filter only terminated employee while grouping them by their gender and reason of termination.
2-4	Generating the bar chart
3	Generating the title for the bar chart

Data Visualization



Observation:

Based on the data visualization above, we could observe that most females are terminated due to retirement and retirement holds the majority reason of why employees are terminated.

This shows that other than age, the company doesn't face any internal issues which causes high layoff and high resignation rate.

Conclusion of Analysis 3: Observation on the gender ratio of the company's terminated employees

After conducting various analysis based on: -

- Observation on the gender ratio of the company's terminated employees
- Observation on the gender ratio of the company's terminated employees based on their gender and age
- Observation on the gender ratio of the company's terminated employees based on their gender, job and age
- Observation on the gender ratio of the company's terminated employees based on their gender and type of termination
- Observation on the gender ratio of the company's terminated employees based on their gender and city
- Observation on the gender ratio of the company's terminated employees based on their gender and department
- Observation on the gender ratio of the company's terminated employees based on their gender and length of service
- Observation on the gender ratio of the company's terminated employees based on their gender and reason

It could be concluded that: -

- Retirement is the majority reason of termination
- Female are more likely to retire as termination
- Female employees are more likely to leave at 13 years of service, whereas male employees are more likely to leave at 8 years of service
- Female employees tends to retire at 24-25 years of service, whereas male employees are more likely to retire at 10-20 years of service
- Most male employees starts to retire at the age of 60 years old, whereas female employees start to retire at the age of 65 years old.
- Older female employee are more likely to be terminated voluntarily, whereas young female worker are more likely to be terminated involuntarily.
- Young and older male worker have the same possibility to be terminated involuntarily, but only older male worker are more likely to be terminated voluntarily.

- Younger employees who worked as cashier, dairy person, shelf socker and baker are more likely to be terminated.
- Older employees from the produce and meat department are more likely to be terminated.

Evidence: -

According to research conducted by the Boston College, they concluded that the amount of time women spend on activities in the sphere of public engagement depends on their ties to their families (BostonCollege, 2013). Relating to the company, we could see that most female employee tends to retire at an old age which shows that they would wish to spend more of their time with families.

As mentioned by wealth awesome, most Canadian choose to retire at the age of 60-70 years old. They also state that by taking the Canada pension plan at the age of 60, employees will receive 36% less compared to age 65years old (WealthAwesome, 2022). This shows why in the company, male employee only leaves at the age of 60 whereas the oldest employee to retire from the company is 65 years old.

Analysis 4: Observation on the company's layoff rate

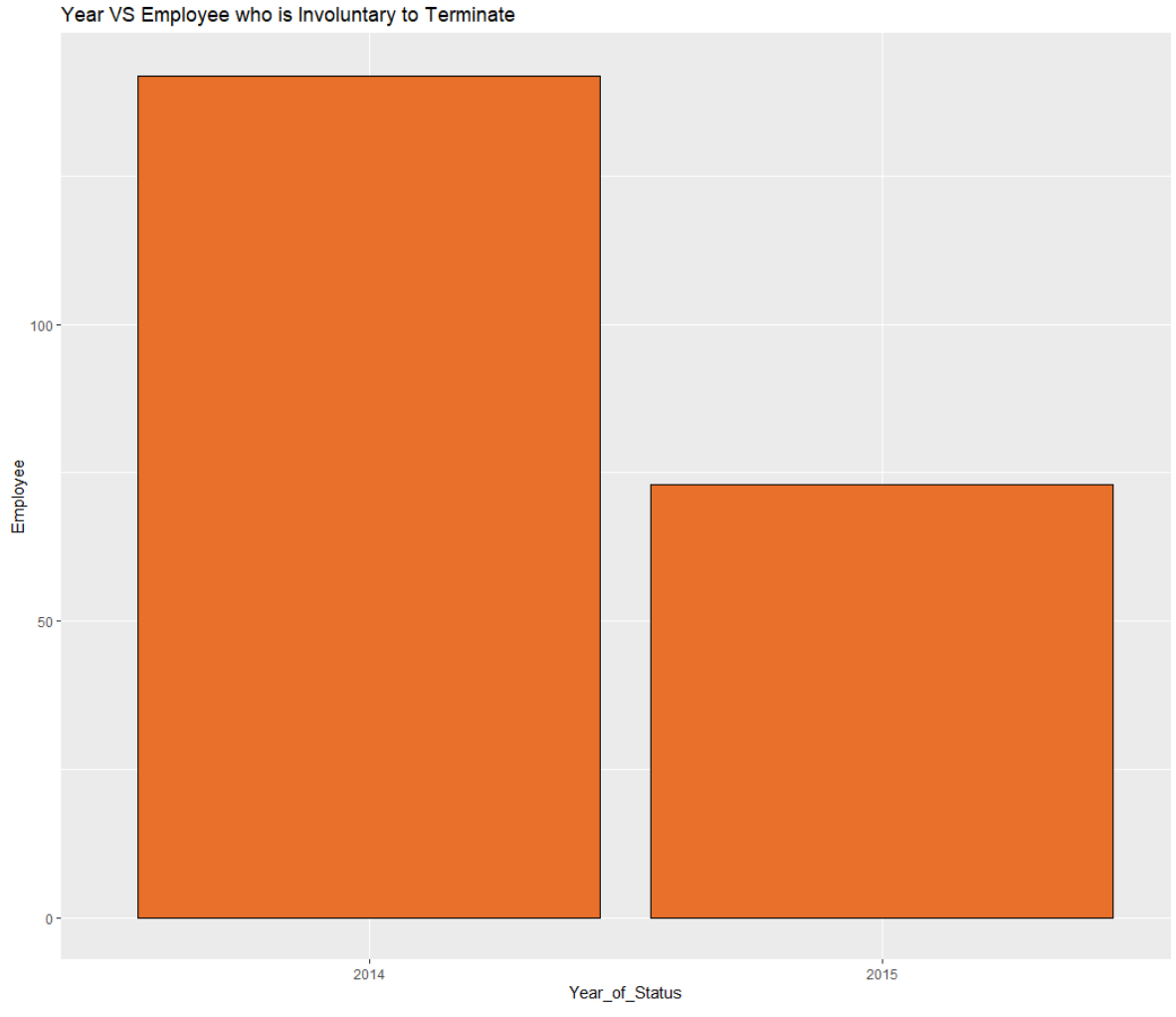
Analysis 4.1: Observation on the company's layoff rate based on each year(most)

Source code

```
#which year has the most?
data2 %>% filter(Status =="TERMINATED" & Type_of_Termination=="Involuntary") %>%
  group_by(Year_of_Status) %>% summarise(Employee =n())%>%
  ggplot(aes(x=Year_of_Status,y=Employee))+geom_bar(stat="identity",fill="#e8702a",col="black")+
  ggtitle("Year VS Employee who is Involuntary to Terminate")
```

line	Explanation
1	To summarize the data status to only filter terminated employee and type of termination to filter involuntary termination
2-3	To generate the bar chart
4	To generate the title for the bar chart

Data Visualization



Observation:

Based on the data visualization above, we could observe that the year 2014 and 2015 have the most involuntary termination by the company with almost 200 employees being terminated in the year 2014 alone.

This shows that 2014 and 2015 have the most layoff termination conducted by the company

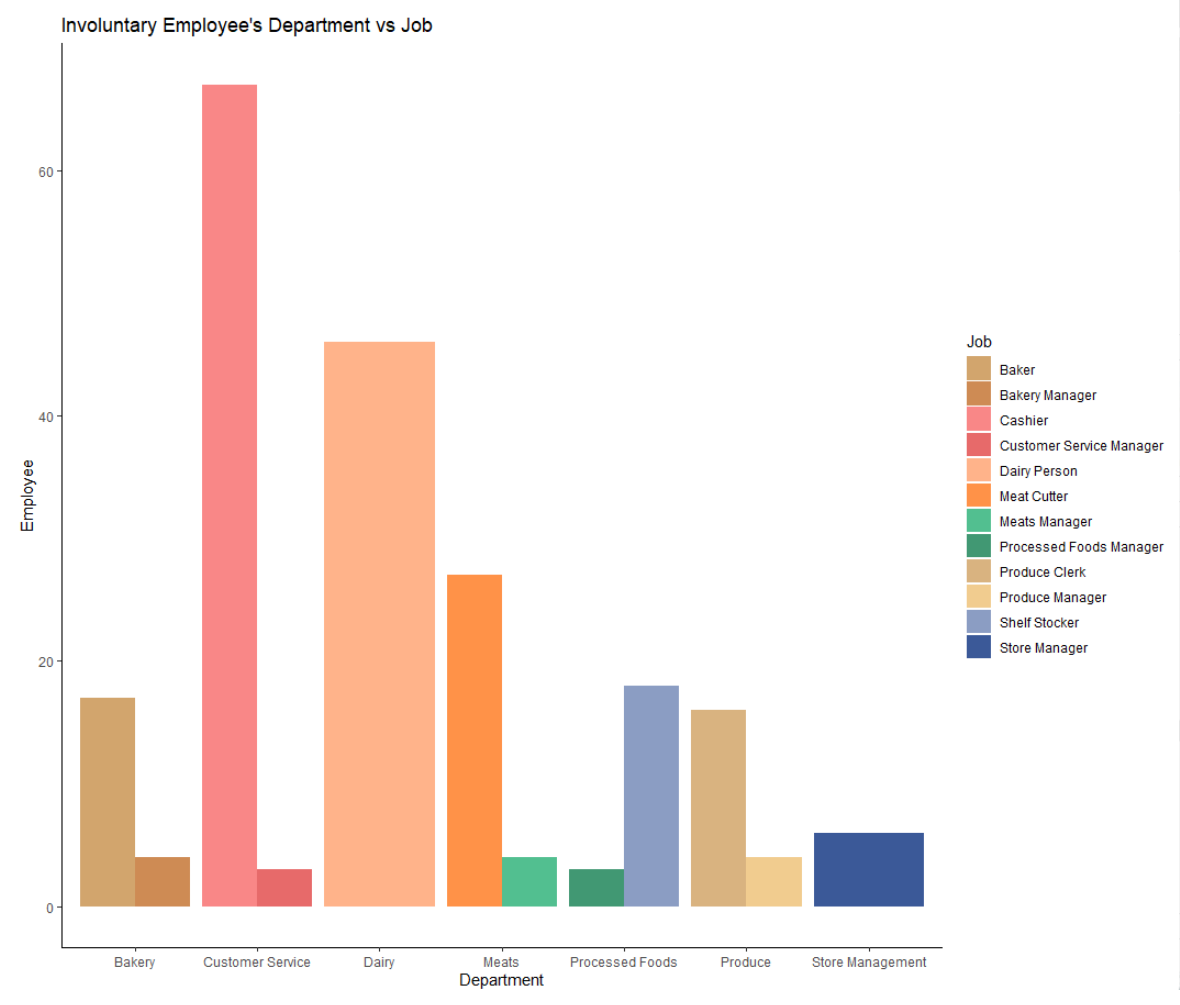
Analysis 4.2: Observation on the company's layoff rate based on their job and department

Source code

```
#job and department
data2 %>% filter(Status =="TERMINATED" & Type_of_Termination=="Involuntary") %>% group_by(Department,Job) %>%
  summarise(Employee =n()) %>% ggplot(aes(x=Department,y=Employee,fill=Job))+
  geom_bar(stat="identity",position ="dodge")+ scale_fill_manual(values=c("#d2a56d","#ce8b54","#f98787",
                                                                          "#e76a6a","#ffb38a","#ff9248","#52bf90",
                                                                          "#419873","#d9b380","#f1cc8f","#8b9dc3",
                                                                          "#3b5998"))+
  ggtitle("Involuntary Employee's Department vs Job")+theme_classic()
```

line	Explanation
1	To summarize the data status to filter only terminated and type of termination to only involuntary terminated employee while group them by their department and job
2-3	Generate bar chart
4	Generate title for the bar chart

Data Visualization



Observation:

Based on the data visualization above, we could observe that most of the involuntary employees are from the customer service department, whereas the dairy department holds the second spot. It is also shown that the amount of supervisors terminated does not affect the amount of employees terminated.

This shows that the cashier department has the highest involuntary termination compared to other departments.

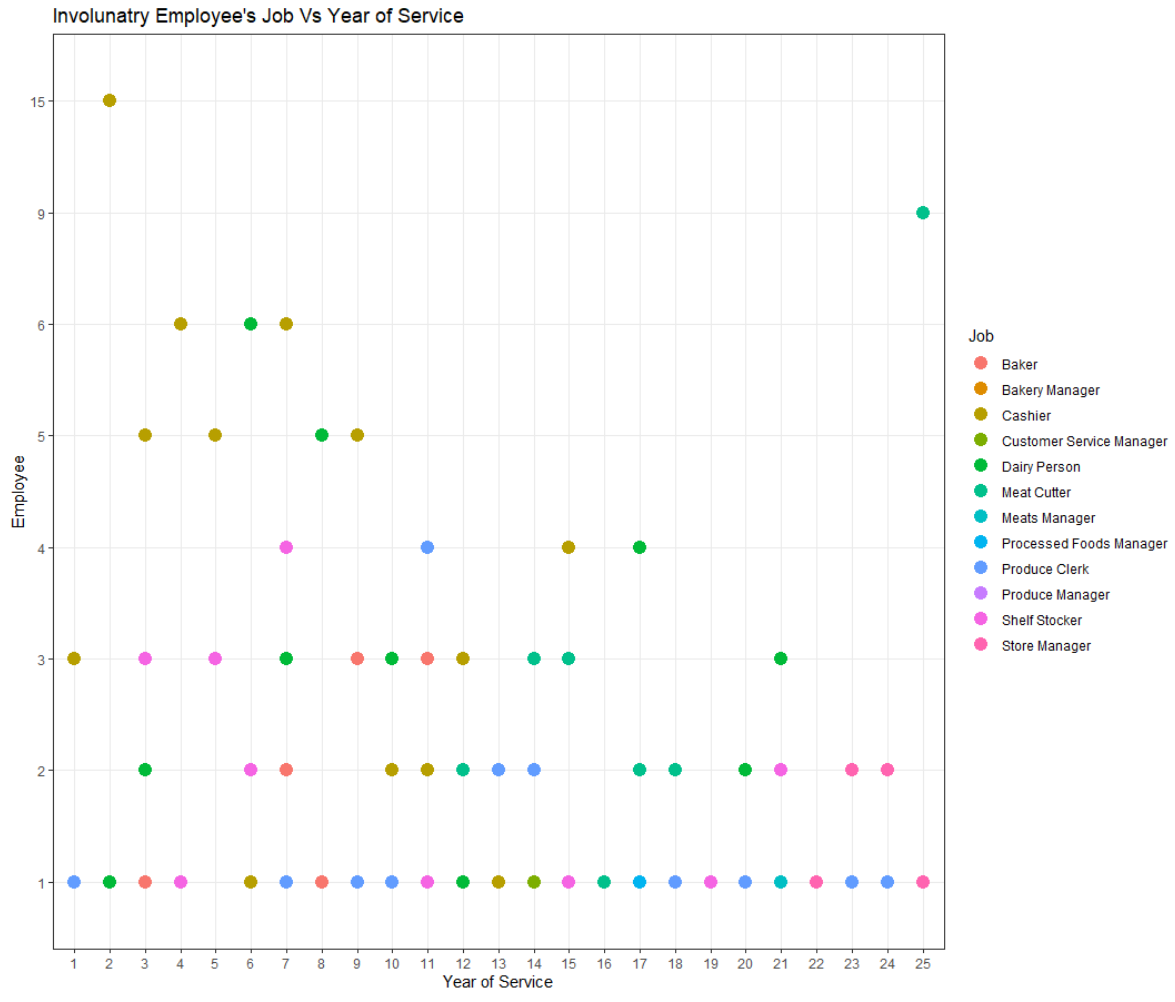
Analysis 4.3: Observation on the company's layoff rate based on their year of service and job

Source code

```
#year of service vs Job
data2 %>% filter(Status =="TERMINATED" & Type_of_Termination=="Involuntary") %>% group_by(Year_of_Service,Job) %>%
  summarise(Employee =n()) %>% ggplot(aes(x=factor(Year_of_Service),y=factor(Employee),col=Job))+
  geom_point(size=4)+theme_bw()+ ggtitle("Involunatry Employee's Job Vs Year of Service")+xlab("Year of Service") +
  ylab("Employee")
```

line	Explanation
1	To summarize the datat status and type of termination to filter only involuntary terminated employee, while group by year of service and job.
2-4	To generate the scatter plot
3	Generate the title for the scatter plot

Data Visualization



Observation:

Based on the data visualization above, we could observe that most terminated cashiers have the lowest year of service, whereas terminated meat cutters have the longest year of service. Other than that, most terminated cashiers have their year of service range between 1 to 12 years whereas meat cutters have their year of service range between 15 to 25 years.

This shows that cashiers are involuntarily terminated at 1 to 12 years of service length, whereas meat cutters are involuntarily terminated at 15 to 25 years of service length.

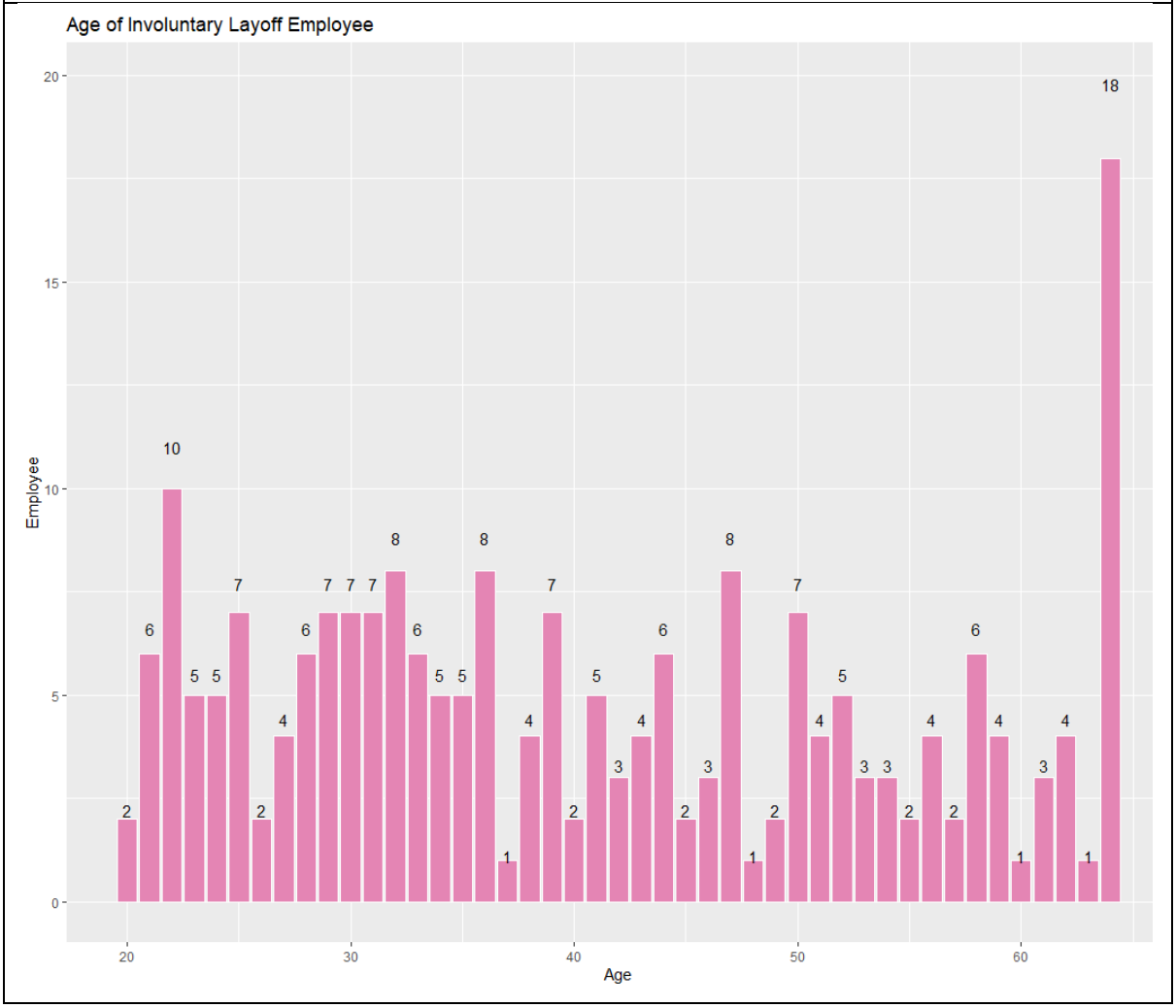
Analysis 4.4: Observation on the company's layoff rate based on their age

Source code

```
#age
data2 %>% filter(Status =="TERMINATED" & Type_of_Termination=="Involuntary" & Reason_of_Termination=="Layoff") %>%
  group_by(Age) %>% summarise(Employee=n()) %>% ggplot(aes(x=Age,y=Employee))+geom_bar(stat="identity",fill="#e485b4",col="white")+
  geom_text(aes(label=Employee),position = position_stack(vjust =1.1))+theme_gray()+
  ggtitle("Age of Involuntary Layoff Employee")
```

line	Explanation
1	To summarize the data status and type of terination to filter only involuntary terminated with layoff as their reason.
2-3	To generate the bar chart
4	To generate the title of the bar chart

Data Visualization



Observation:

Based on the data visualization above, we could observe that most of the involuntary terminated employees are aged above 60. Other than that, there are a huge amount of involuntary terminated employees that are aged between 20 to 30 years old.

This shows that most of the involuntary terminated employees either are terminated between 20 to 30 years old, or they are terminated at the age of 60 and above.

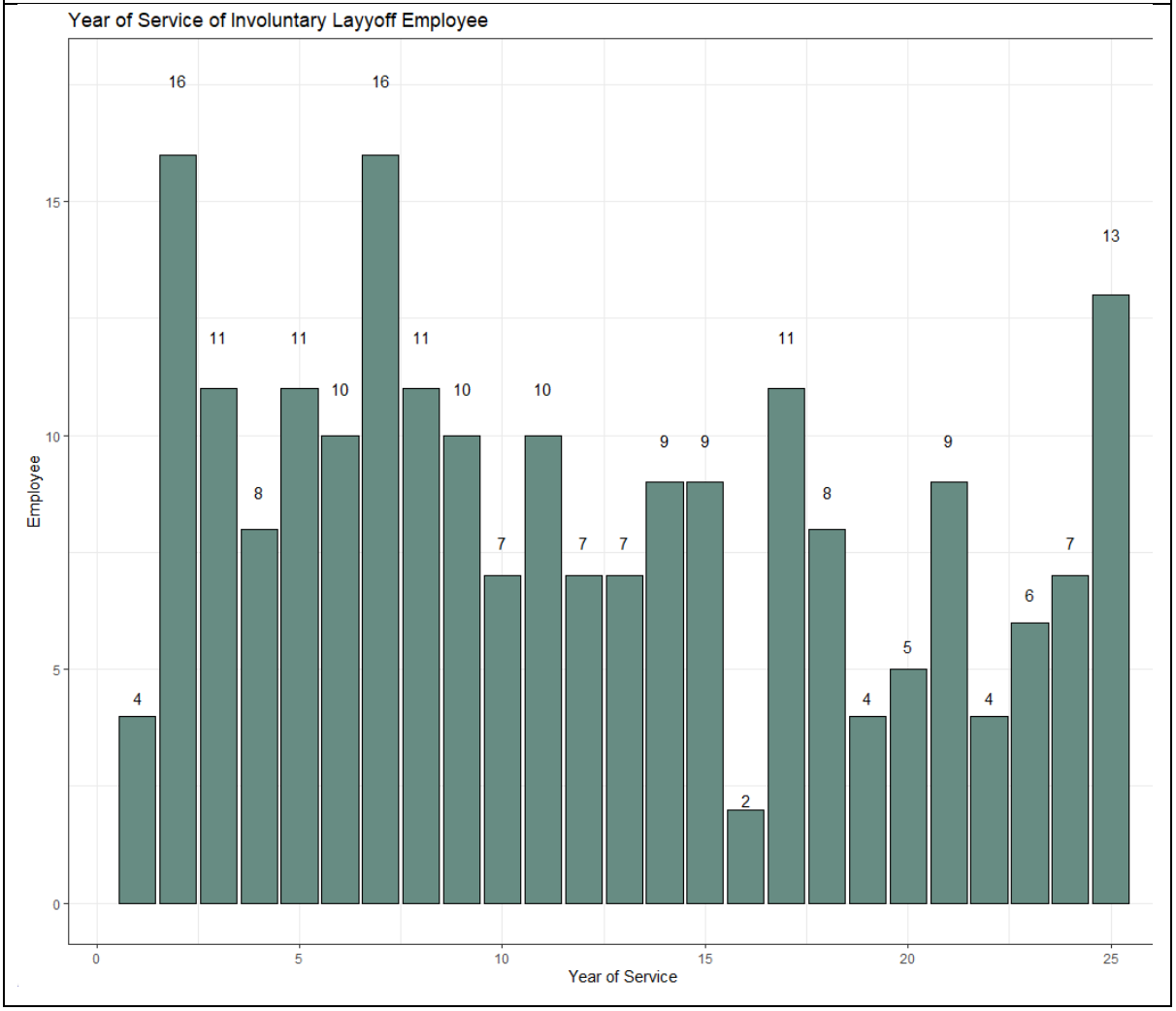
Analysis 4.5: Observation on the company's layoff rate based on their length of service

Source code

```
#length
data2 %>% filter(Status =="TERMINATED" & Type_of_Termination=="Involuntary" & Reason_of_Termination=="Layoff") %>%
  group_by(Year_of_Service) %>% summarise(Employee=n()) %>% ggplot(aes(x=Year_of_Service,y=Employee))+
  geom_bar(stat="identity",fill ="#678c83",col="black")+xlab("Year of Service")+
  geom_text(aes(label=Employee),position = position_stack(vjust =1.1))+theme_bw()+
  ggtitle("Year of Service of Involuntary Layyoff Employee")
```

line	Explanation
1	To summarize the data status, type of termination and reason of termination to filter to employee who are involuntary terminated with layoff as ther reason
2-4	To generate the bar chart
5	To generate the title for the bar chart

Data Visualization



Observation:

Based on the data visualization above, we could observe that most involuntary terminated employees are terminated with a year of service of 0 to 10 years. There are lesser involuntary terminated employees terminated at the year of service of 15 to 20 years. But there is an increase in involuntary termination for employees with 25 year of service.

This shows that most involuntary terminated employees are employees with 0 to 10 years of service and 25 years of service.

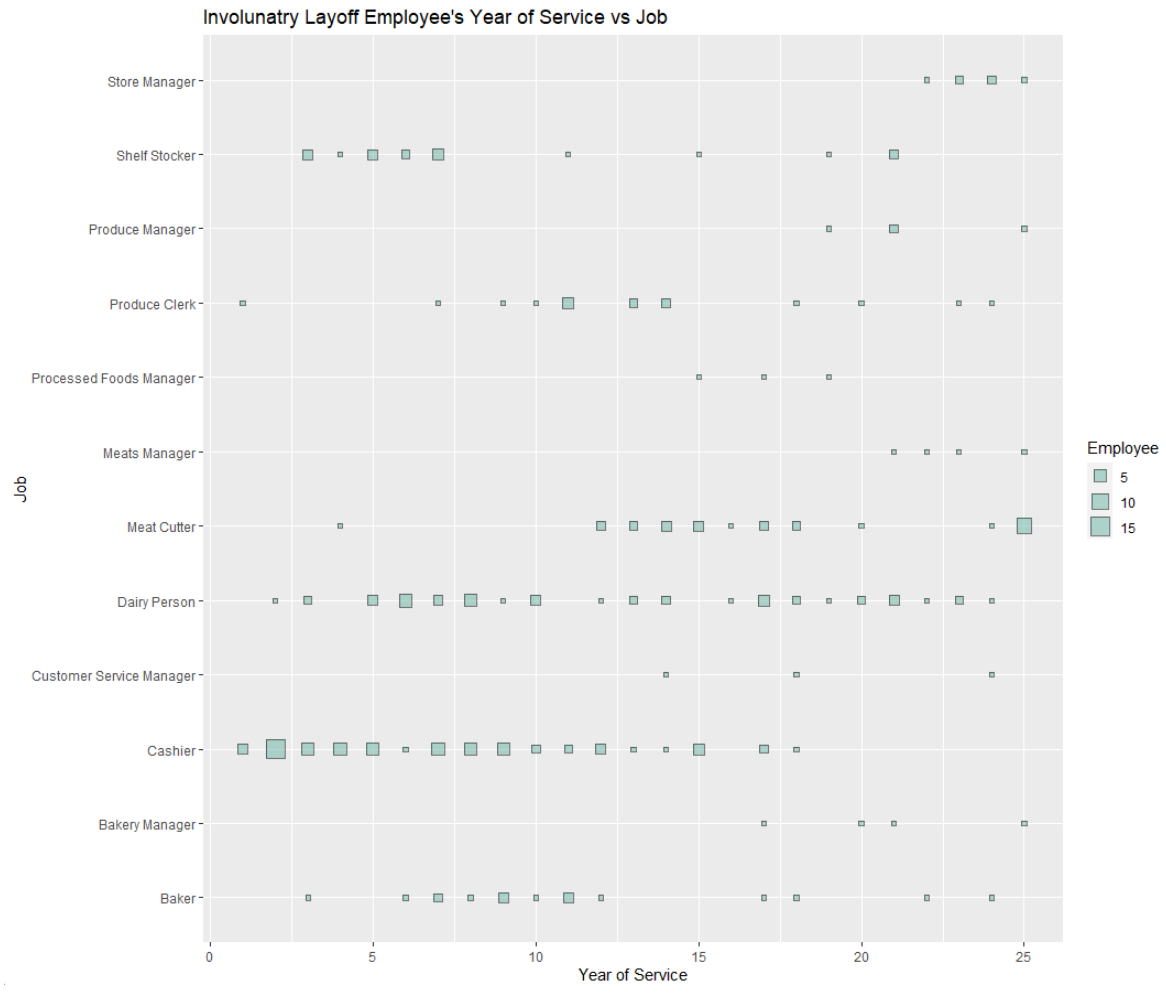
Analysis 4.6: Observation on the company's layoff rate based on their job and their year of service

Source code

```
#job
data2 %>% filter(Status =="TERMINATED" & Type_of_Termination=="Involuntary" & Reason_of_Termination=="Layoff") %>%
  group_by(Year_of_Service,Job) %>% summarise(Employee=n()) %>% ggplot(aes(x=Job,y=Year_of_Service,size=Employee))+
  geom_point(color="black",fill="#69b3a2",shape=22,alpha=0.5,stroke=1)+coord_flip()+ylab("Year of Service")+
  ggtitle("Involunatry Layoff Employee's Year of Service vs Job")
```

line	Explanation
1-2	To summarize the data status, type of termination and reason of termination to filter only involuntary terminated employee with reason of layoff while grouping them with their year of service.
2-3	To generate the scatter plot graph
4	To generate the title for the scatter plot

Data Visualization



Observation:

Based on the data visualization above, we could observe that involuntary terminated shelf stocker, dairy person, cashier and baker have years of service of 0 to 10 years of service, whereas meat cutter and produce clerk have years of service of 10 to 20 years of service. Most involuntary terminated bakery managers are terminated with a length of service of 17 to 25 years and involuntary terminated processed food managers are terminated with length of service 15 to 20 years old.

This shows that young employees who work as shelf stocker, dairy person, cashier and baker are more likely to be involuntarily terminated, whereas employees who worked as meat cutter and produce clerk with a 10 to 20 years of service are more likely to be involuntarily terminated.

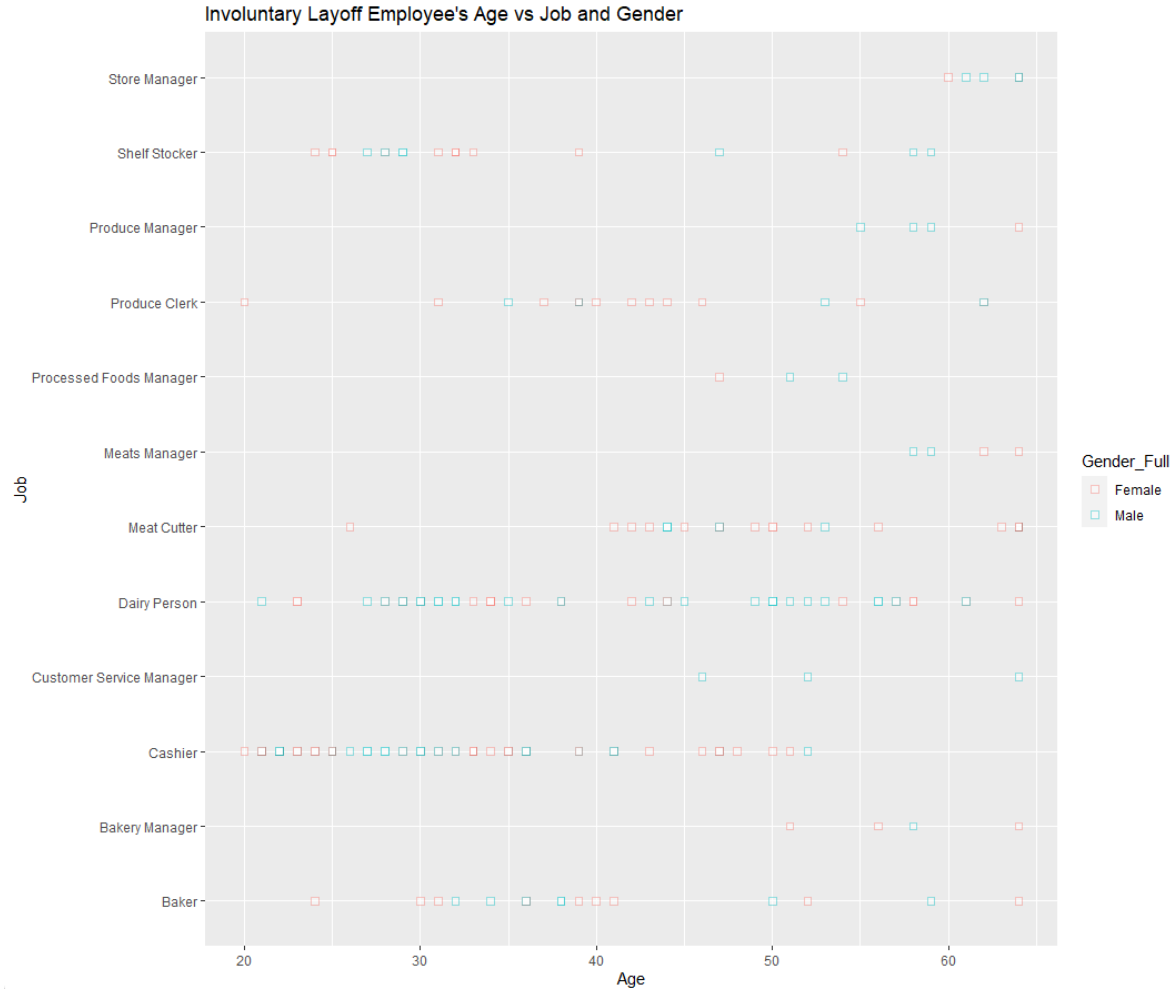
Analysis 4.7: Observation on the company's layoff rate based on their gender, age and job

Source code

```
#gender age job
data2 %>% filter(Status =="TERMINATED" & Type_of_Termination=="Involuntary" & Reason_of_Termination=="Layoff") %>%
  group_by(Age,Job,Gender_Full) %>% ggplot(aes(x=Age,y=Job,col=Gender_Full))+
  geom_point(shape=22,fill="white",alpha=0.4,stroke=1,size=2)+ggtitle("Involuntary Layoff Employee's Age vs Job and Gender")
```

line	Explanation
1	To summarize the data status, type of termination and reason of termination to filter only involuntary terminated employee with reason of layoff.
2-3	To generate the scatter plot graph
3	To generate the title for the scatter plot

Data Visualization



Observation:

Based on the data visualization above, we could observe that most dairy persons and cashiers who are terminated involuntarily are male who are between 20 to 35 years old. Most produce clerks who are involuntarily terminated at the age of 30 to 45 years old are mostly women.

Conclusion of Analysis 4: Observation on the company's layoff rate

After conducting various analysis based on: -

- Observation on the company's layoff rate based on each year(most)
- Observation on the company's layoff rate based on their job and department
- Observation on the company's layoff rate based on their year of service and job
- Observation on the company's layoff rate based on their age
- Observation on the company's layoff rate based on their length of service
- Observation on the company's layoff rate based on their job and their year of service
- Observation on the company's layoff rate based on their gender, age, and job

It could be concluded that: -

- The year 2014 and 2015 have the highest involuntary termination rate.
- Customer service and dairy department have the most involuntary teminated employees
- Most involuntary terminated employees are aged between 20 to 30 years old and 60 and above.
- Most involuntary terminated employees are terminated with 0 to 10, and 25 and above length of service
- Young employees who worked as shelf stocker, dairy person, cashier and baker are more likely to be terminated involuntarily
- Old employees who worked as meat cutter and produce clerk are more likely to be terminated involuntarily
- Involuntary termination of dairy person and casheri are mostly male who aged 20 to 35 years old
- Involuntary termination of produce clerk are mostly female who aged 30 to 45 years old.

Evidence: -

As mentioned by AlignToughts, reasons why younger employees are terminated involuntarily is due to employee misconduct, poor performance, absenteeism, ethical issues, violations of the company policies etc. Relating back to the company, most young employees are terminated are due to poor performance whereas mentioned before, the company have high termination rate at the month December which indicates most employee review are conducted at December.

As mentioned by Samifiru Tumarkin Lpp, the older an employee is, the longer they have spent in a company's employ, the more severance they will be entitled to (Simafiru, 2017). Relate back to the company, the reason why company tends to terminate younger employee is due to compensation of terminating an older employee.

As mentioned by SHRM, most termination of older employees are due to lack of desired skills and not age bias (SHRM,2017).Relating back to the company, reason why older employees are terminated involuntarily is due to outdated skillset which they bring in the workforce.

Analysis 5: Observation on the voluntary termination rate of the company

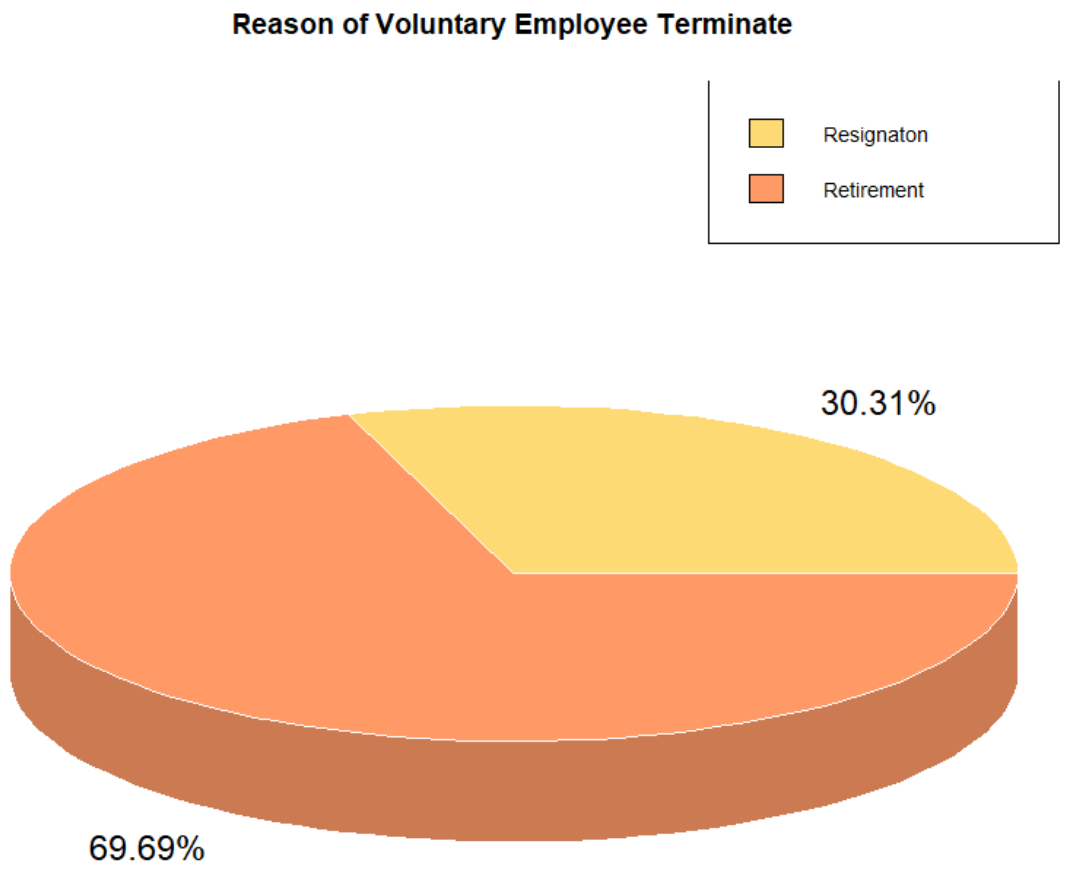
Analysis 5.1: Observation on the voluntary termination rate of the company based on terminated employee's reason

Source code

```
#reason
r_volun= table(data2 %>% filter(Status =="TERMINATED" & Type_of_Termination=="Voluntary") %>% select(Reason_of_Termination))
rVPercent = round(100*r_volun/sum(r_volun),2)
pie3D(r_volun,labels=paste0(rVPercent,"%"),cex=0.7,radius=1,main="Reason of Voluntary Employee Terminate",border="white", col=c("#feda75","#ff9966"))
legend("topright",c("Resignaton","Retirement"),cex=0.9,fill=c("#feda75","#ff9966"))
```

line	Explanation
1	The variable r_volun is assign to select and store status which filter terminated employee and type of termination which filter voluntary using the table function while selecitng the reason of termination.
2	To generate a 3D pie chart
3	To generate legend to show information of the pie chart

Data Visualization



Observation:

Based on the data visualization above, we could observe that most of the voluntary terminated employees are due to retirement standing with 69.69% and resignation which hold 30.31%

This shows that most voluntary terminated employees are terminated due to retirement.

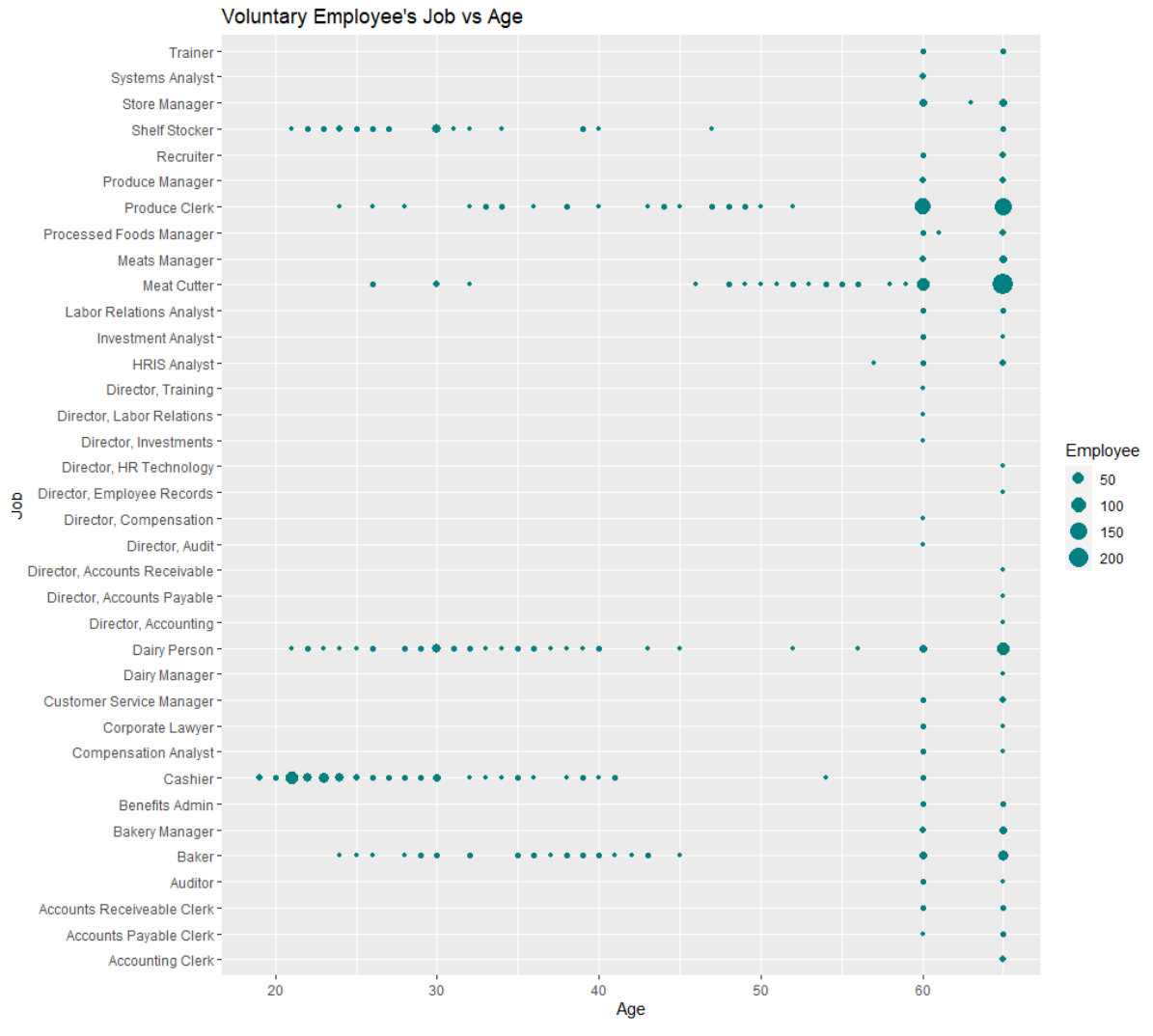
Analysis 5.2: Observation on the voluntary termination rate of the company based on terminated employee's age and job

Source code

```
#age + job
data2 %>% filter(Status =="TERMINATED" & Type_of_Termination=="Voluntary") %>% group_by(Age,Job) %>%
  summarise(Employee =n()) %>% ggplot(aes(x=Age,y=Job,size=Employee))+geom_point(col="#008080")+
  ggtitle("Voluntary Employee's Job vs Age")
```

line	Explanation
1	To summarize the data status and type of termination to filter only voluntary terminated emplloyee while grouping them by their age and job
2	To generate the scatter plot
3	To generate the title for the scatter plot

Data Visualization



Observation:

Based on the data visualization above, we could observe that most dairy people, baker, shelf stocker and cashier who are terminated voluntarily are aged between 20 to 40 years old. Most produce clerks who are voluntarily terminated at the age of 30 to 50 years old. Most meat cutters who are voluntarily terminated are between the age of 45 to 60 years old.

This shows that younger employees tend to resign from jobs such as dairy person, baker, shelf stocker and cashier, and older employees tend to resign from jobs such as meat cutter and produce clerk.

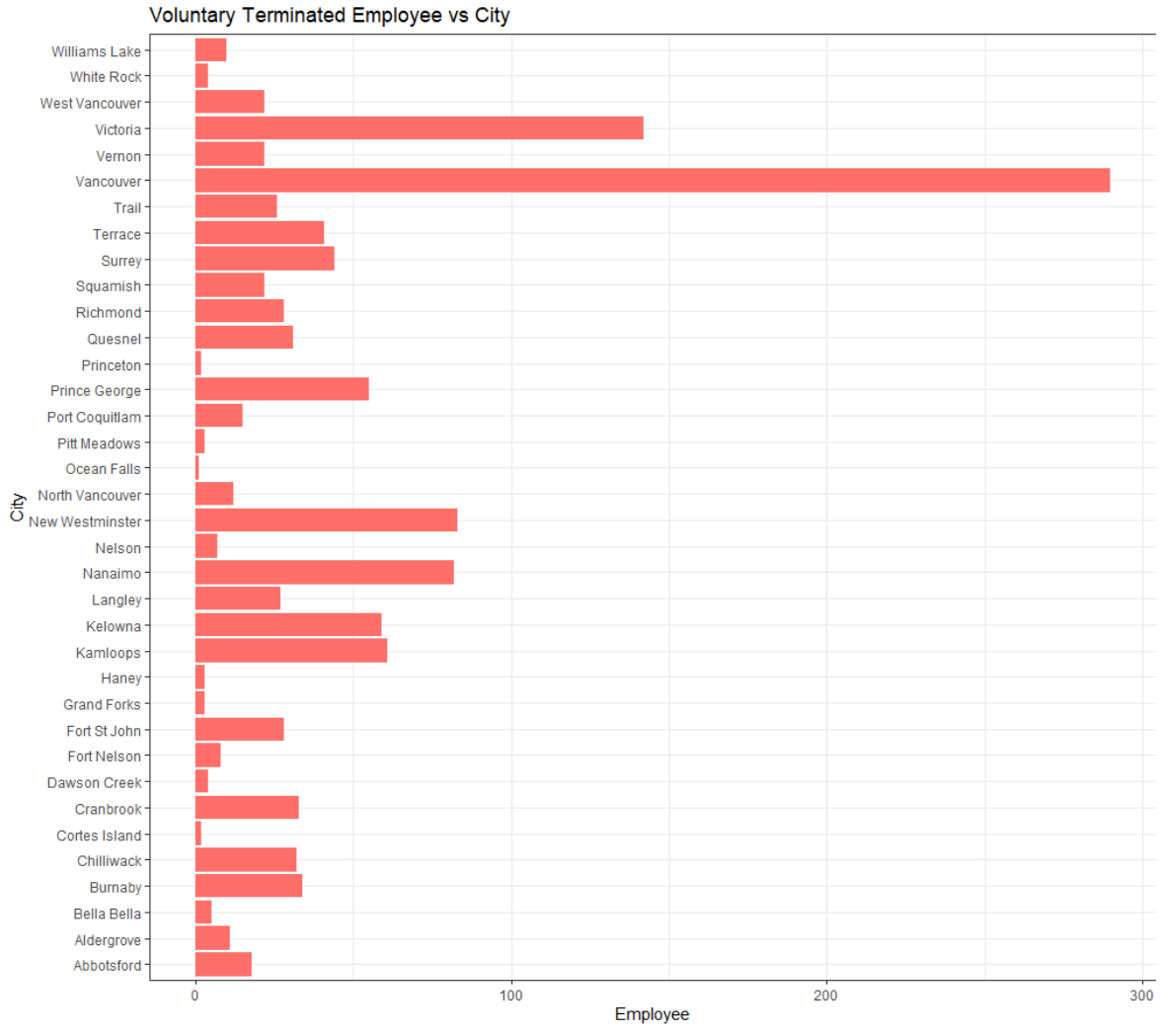
Analysis 5.3: Observation on the voluntary termination rate of the company based on terminated employee's city

Source code

```
#city
data2 %>% filter(Status =="TERMINATED" & Type_of_Termination=="Voluntary") %>% group_by(City) %>% summarise(Employee =n()) %>%
  ggplot(aes(x=City,y=Employee))+geom_bar(stat="identity",fill="#ff6f69")+coord_flip()+ggtitle("Voluntary Terminated Employee vs City")+
  theme_bw()
```

line	Explanation
1	To summarize the data status and type of termination to voluntary terminated employee while grouping them by city
2-3	Generate the bar chart while creating the title for the abr chart

Data Visualization



Observation:

Based on the data visualization above, we could observe that the country with the most voluntary terminated employees are from vancouver. Other than that, cities like victoria and new Westminster also have lots of employees who terminated voluntarily.

This shows that major cities like vancouver, victoria and new Westminster have lots of voluntary terminated employees.

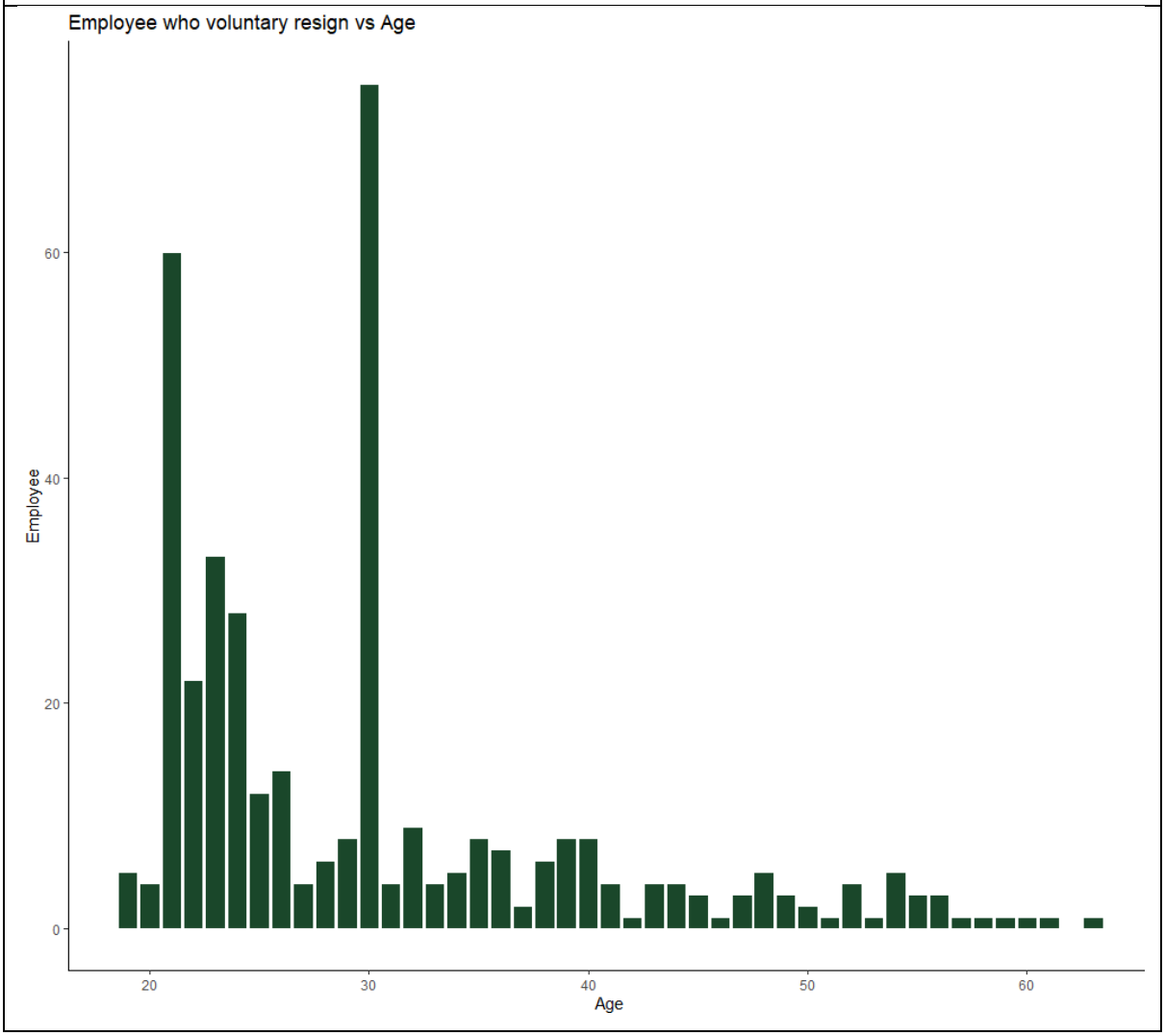
Analysis 5.4: Observation on the voluntary termination rate of the company based on resigned employee's age

Source code

```
#resign()
#resign + age
data2 %>% filter(Status =="TERMINATED" & Type_of_Termination=="Voluntary" & Reason_of_Termination=="Resignaton") %>% group_by(Age) %>%
  summarise(Employee=n()) %>% ggplot(aes(x=Age,y=Employee))+geom_bar(stat="identity",fill="#1a472a",col="white")+theme_classic()+
  ggtitle("Employee who voluntary resign vs Age")
```

line	Explanation
1	To summarize the ata status, type of termination and reason of termination to filter only voluntary terminaed employee with the reason of resignation while group them by their age
2	To generate the bar chart
3	To generate the title for the bar chart

Data Visualization



Observation:

Based on the data visualization above, we could observe that most of the voluntary terminated employees are aged 20 to 30 years old. Other than that, there is a decreasing amount in voluntary termination after the age of 40.

This shows that most employees choose to resign at the age of 20 to 30 years old.

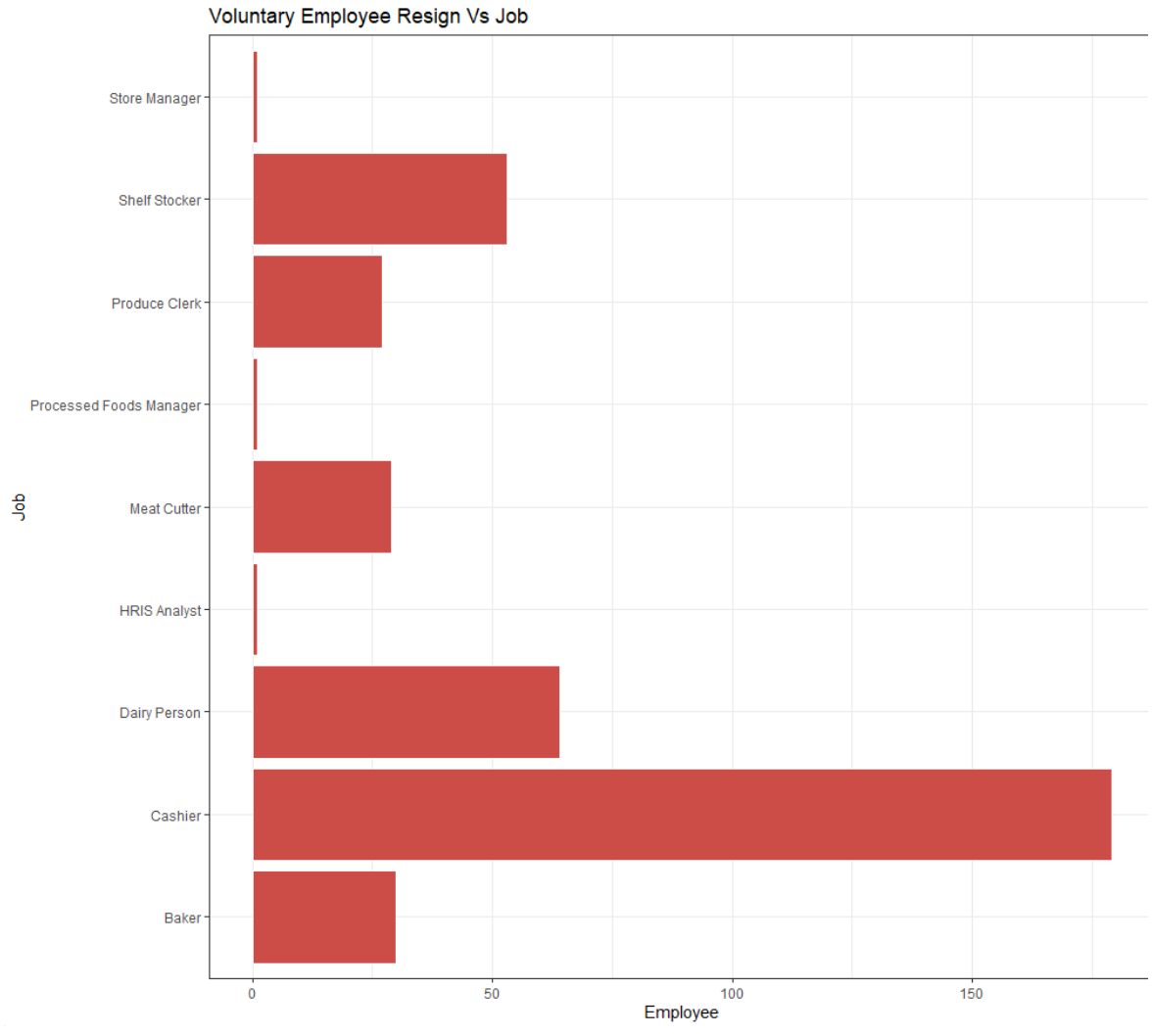
Analysis 5.5: Observation on the voluntary termination rate of the company based on resigned employee's job

Source code

```
#resign vs job
data2 %>% filter(Status =="TERMINATED" & Type_of_Termination=="Voluntary" & Reason_of_Termination=="Resignaton") %>%
  group_by(Job) %>% summarise(Employee=n()) %>% ggplot(aes(x=Job,y=Employee))+
  geom_bar(stat="identity",fill="#cc4d48",col="white")+coord_flip()+
  ggtitle("Voluntary Employee Resign Vs Job")+theme_bw()
```

line	Explanation
1	To summarize the data status, type of termination and reason of termination to filter only voluntary terminated emlpoyee with resignation as reason while group them by their job.
2-3	To generate the bar chart
4	To generate the title for the bar chaart

Data Visualization



Observation:

Based on the data visualization above, we could observe that most employees choose to resign from being a cashier. Other than that, jobs like baker, dairy person, meat cutter, produce clerk and shelf stocker too have a high amount of voluntarily terminated employees.

This shows that jobs such as baker, cashier, meat cutter, produce clerk and shelf stocker have poor employee retention rate.

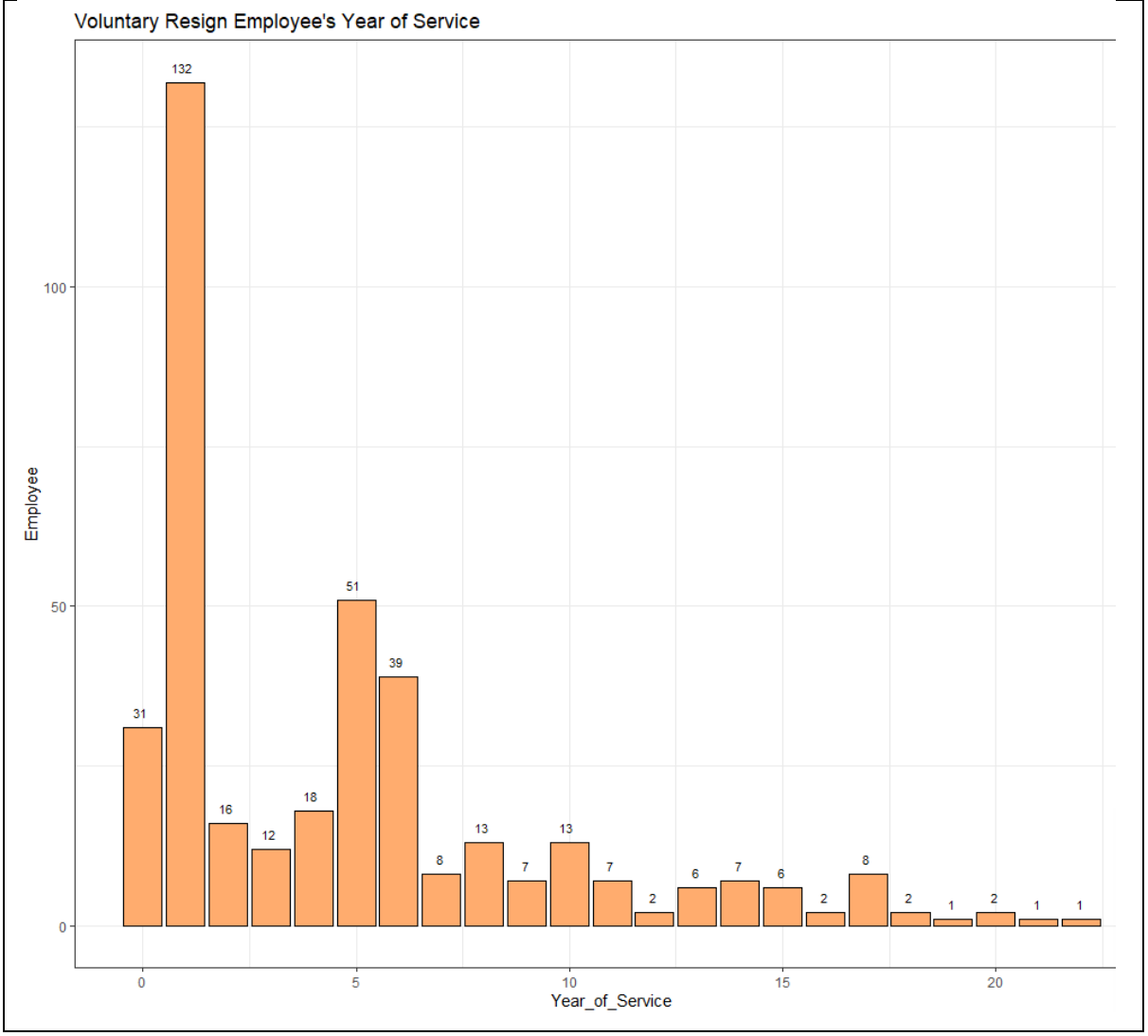
Analysis 5.6: Observation on the voluntary termination rate of the company based on resigned employee's length of service

Source code

```
#resign vs length
data2 %>% filter(Status =="TERMINATED" & Type_of_Termination=="Voluntary" & Reason_of_Termination=="Resignaton") %>%
  group_by(Year_of_Service) %>% summarise(Employee=n()) %>%
  ggplot(aes(x=Year_of_Service,y=Employee))+geom_bar(stat="identity",fill="#ffac6e",col="black")+
  geom_text(aes(label=Employee),hjust=0.7,vjust=-1,col="black",size=3)+theme_bw()+
  ggtitle("Voluntary Resign Employee's Year of Service")
```

line	Explanation
1-2	To summarize the data status, type of termination and reason of termination to filter only voluntary terminated employee with resignation as reason while group them by their year of service
3-4	To generate the bar chart
5	To generate title for the bar chart

Data Visualization



Observation:

Based on the data visualization above, we could observe that most employees who resign are employees with 0 to 2 years of length of service. Moreover, there are a lot of employees who resign after 5 years of service with the company.

This shows that most resign employees are employees who joined the workforce and employees with 5 years of service with the company.

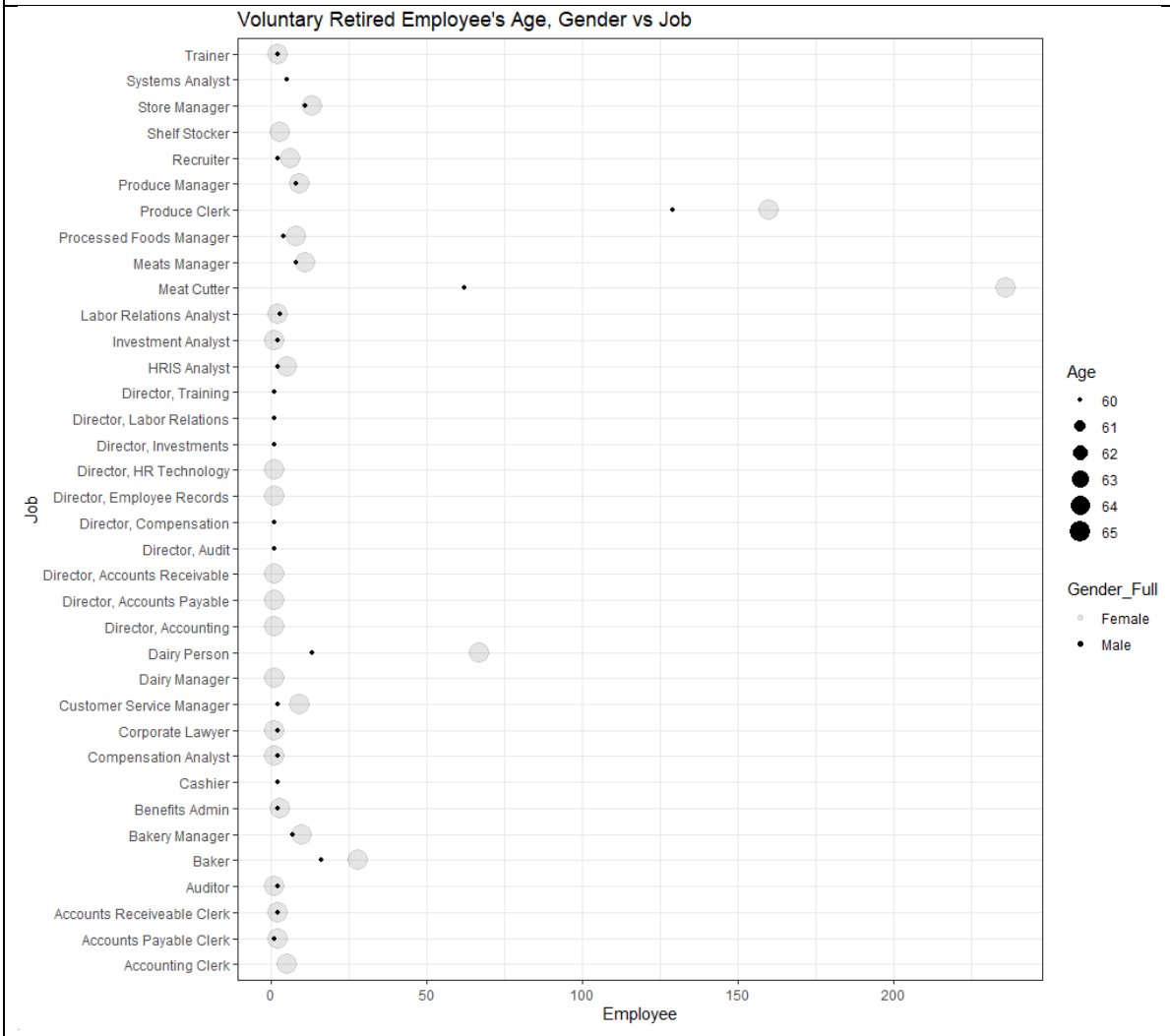
Analysis 5.7: Observation on the voluntary termination rate of the company based on retired employee's job, gender, and age

Source code

```
#retire()
#retire: job vs gender vs age
data2 %>% filter(Status =="TERMINATED" & Type_of_Termination=="Voluntary" & Reason_of_Termination=="Retirement") %>%
  group_by(Age,Gender_Full,Job) %>% summarise(Employee=n()) %>%
  ggplot(aes(x=Job,size=Age,y=Employee,alpha=Gender_Full))+geom_point()+coord_flip()+
  ggtitle("Voluntary Retired Employee's Age, Gender vs Job")+theme_bw()
```

line	Explanation
1-2	To summarize data status, type of termination and reason of termination to filter only voluntary terminated employee with retirement as reason while grouping them with their age, gender and job.
3	To generate the scatter plot graph
4	To generate title for the scatter plot

Data Visualization



Observation:

Based on the data visualization above, we could observe that most male employees from different jobs choose to resign at a young age. Whereas female employees choose to resign at an older age. Females from jobs like Produce clerk, meat cutter and dairy cutter have a lot of female employees resign.

This shows that female employees resign later than male employees. Moreover, jobs like produce clerk, meat cutter and dairy person have a lot of resignation.

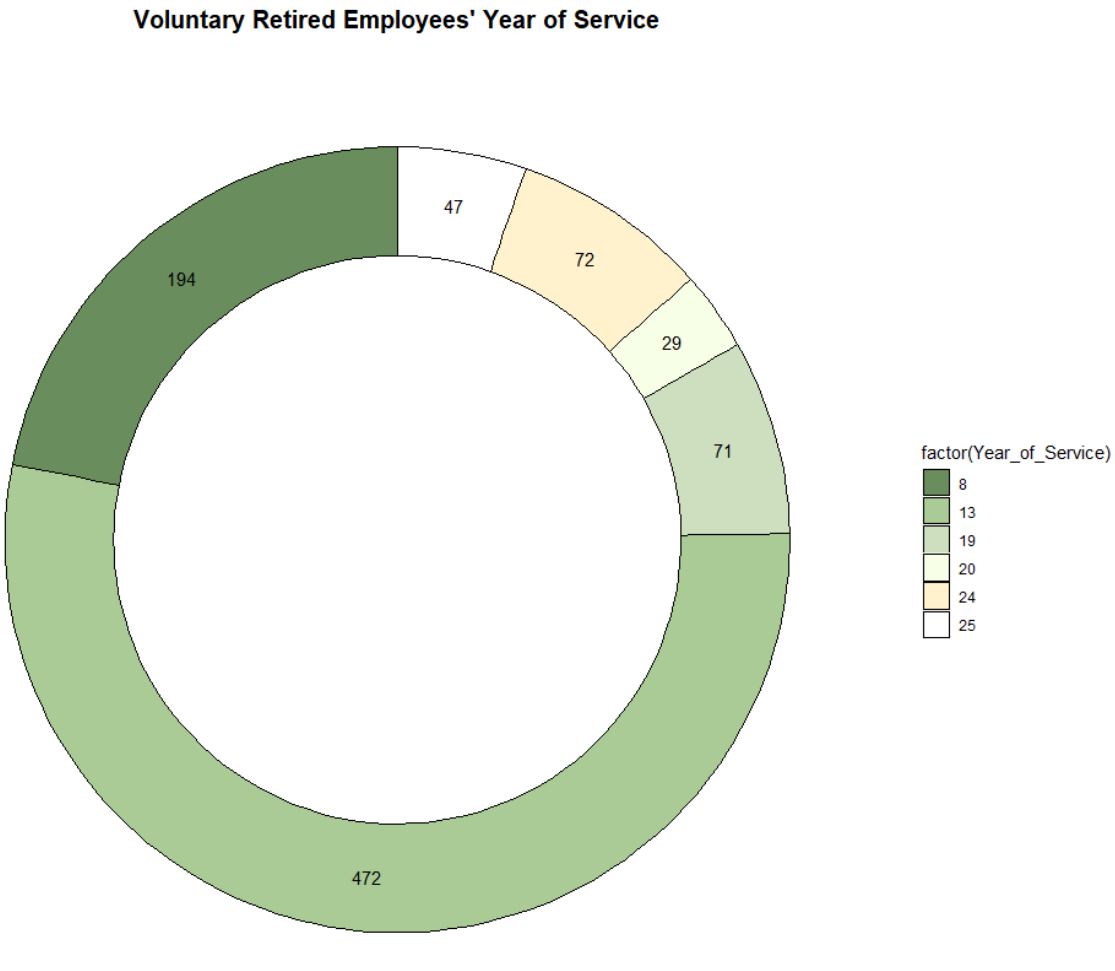
Analysis 5.8: Observation on the voluntary termination rate of the company based on retired employee's length of service

Source code

```
#length
data2 %>% filter(Status =="TERMINATED" & Type_of_Termination=="Voluntary" & Reason_of_Termination=="Retirement") %>%
  group_by(Year_of_Service) %>% summarise(Employee=n()) %>% ggplot(aes(x=3,y=Employee,fill=factor(Year_of_Service)))+
  geom_col(col="black")+coord_polar(theta="y")+ xlim(c(0.2,3.5))+theme(panel.background = element_rect(fill = "white"),
                      axis.title = element_blank(),axis.ticks = element_blank(),axis.text = element_blank())+
  geom_text(aes(label=Employee),position = position_stack(vjust =0.5))+
  ggtitle("Voluntary Retired Employees' Year of Service")+
  scale_fill_manual(values=c("#698d5d","#aacb96","#cddfbf","#f8ffe7","#fff2cc","white"))+
  theme(plot.title=element_text(size=15,face="bold",hjust = 0.5))
```

line	Explanation
1-2	Summarized the data of status which filters terminated employees with the type of termination voluntary with termination reason retirement while group by the year of service
3-7	Generate donut chart
4	Add legend to help identify information of donut chart
5	Add title for the donut chart

Data Visualization



Observation:

Based on the data visualization above, we could observe that most voluntary terminated employees are terminated with the year of service of 13 years. There are lesser voluntary terminated employees terminated at the year of service of 25 years. But there is an increase in voluntary termination for employees with 8 year of service.

This shows that most voluntary terminations are employees with 8years of service and 13 years of service.

Conclusion of Analysis 5: Observation on the voluntary termination rate of the company

After conducting various analysis based on: -

- Observation on the voluntary termination rate of the company based on terminated employee's reason
- Observation on the voluntary termination rate of the company based on terminated employee's age and job
- Observation on the voluntary termination rate of the company based on terminated employee's city
- Observation on the voluntary termination rate of the company based on resigned employee's age
- Observation on the voluntary termination rate of the company based on resigned employee's job
- Observation on the voluntary termination rate of the company based on resigned employee's length of service
- Observation on the voluntary termination rate of the company based on retired employee's job, gender, and age
- Observation on the voluntary termination rate of the company based on retired employee's length of service

It could be concluded that: -

- Retirement is the higgest reason why employees are terminated voluntarily.
- Most young employees chooose to be terminated voluntarily are from jobs such as dairy person, baker, shelf stocker and cashier.
- Most senior employees choose to be terminated voluntarily are ffrom jobs like meat cutter and produce clerk.
- Major cities like vancouver, victoria and west minster have the most voluntary termination of employees.
- Most employees choose to resign at the age of 20 to 30 years old
- Jobs like baker, cashier, meat cutter, produuce clerk and shelf stocker have the most resign employees
- Most resign employees have 0 to 2 years and 5 years of service with the company
- Male employees tend to resign earlier than female employees

- Most voluntary terminated employees are temrinated with 8 and 13 years of service with the company.

Evidence: -

It is mentioned by statistics Canada, the number of stay-at-home mother with at least one child under 16 are 444700(Statistic Canada, 2019). This shows that most of the female who resign at the age of 20 to 30 years old maybe due to starting an new family which requires them to stay at home to cater.

It is mentioned by TIME, most of the young employees who choose to resign is due to burnout, poor treatment and low income(TIME, 2021). Relating back to the company, we could see a high resignation rate for jobs like baker, cashier, meat cutter, produuce clerk and shelf stocker which proven to be tiring and low paying.

Analysis 6: Observation on the company's hiring interest throughout the year

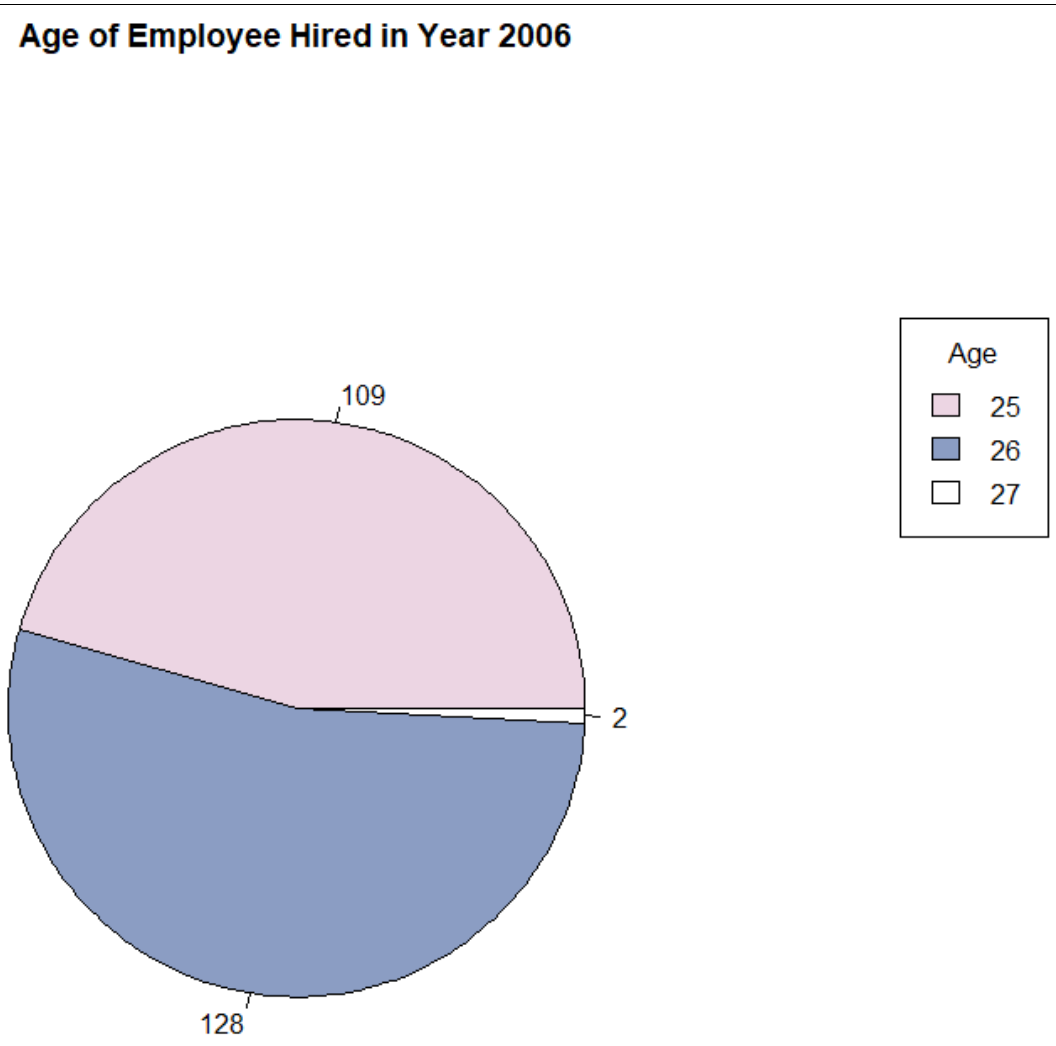
Analysis 6.1: Observation on the company's hiring interest throughout the year based on employees age in 2006

Source code

```
#age
#in 2006
hire_2006 = table(data2 %>% filter(Year_of_Status=="2006",Hire_Year=="2006") %>% select(Age))
pie(hire_2006,label=hire_2006,main="Age of Employee Hired in Year 2006",col=c("#ecd5e3","#8b9dc3","white"),border="black")
legend("topright",c("25","26","27"), cex =1, fill=c("#ecd5e3","#8b9dc3","white"),title="Age")
```

line	Explanation
1	The variable hire_2006 is assign to select and store year of status which filter the year 2006 and hire year which filter the year 2006 using the table function while selecitng the age.
2	To generate a pie chart
3	To generate legend to show information of the pie chart

Data Visualization



Observation:

Based on the data visualization above, we could observe that during 2006, the ages of hired employees are between 25 to 27 years old where most employees are hired at the age of 26.

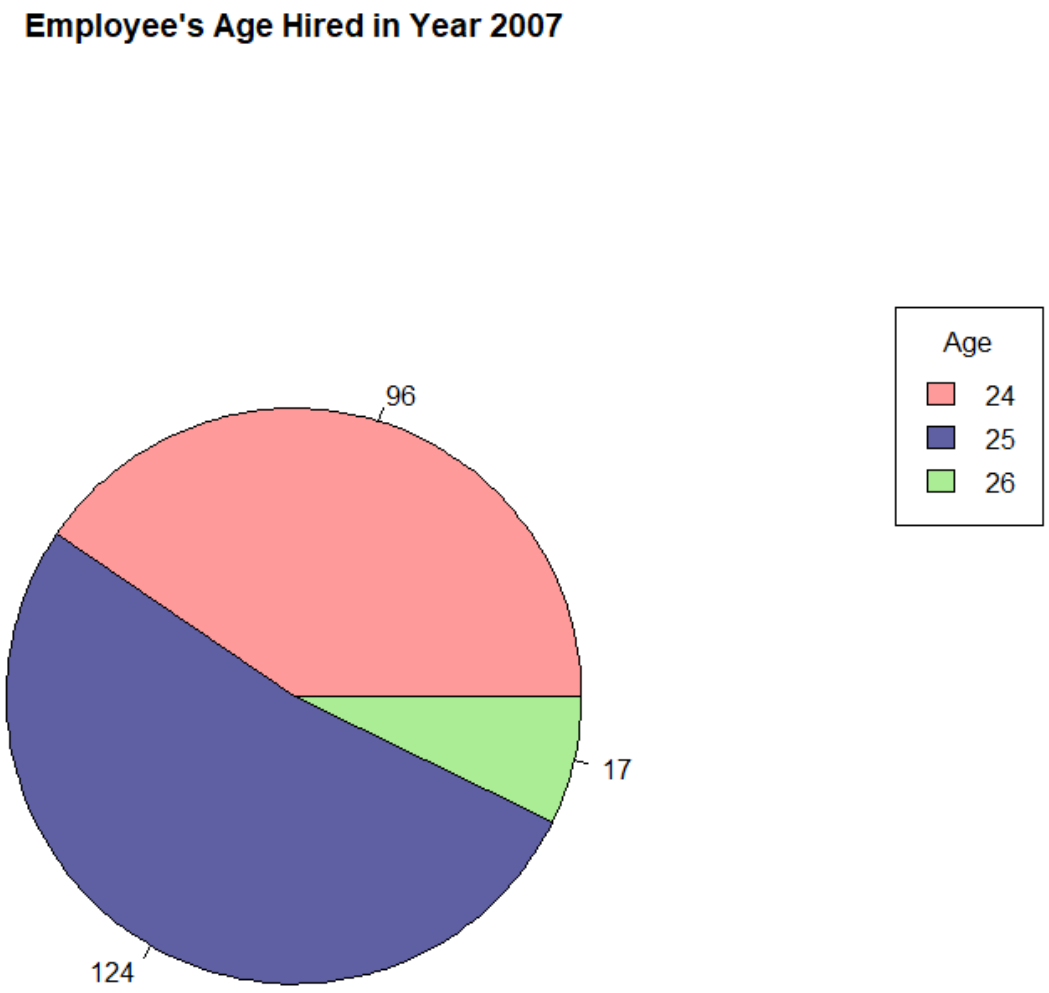
Analysis 6.2: Observation on the company's hiring interest throughout the year based on employees age in 2007

Source code

```
#in 2007
hire_2007 = table(data2 %>% filter(Year_of_Status=="2007",Hire_Year=="2007") %>% select(Age))
pie(hire_2007,label=hire_2007,main="Employee's Age Hired in Year 2007",col=c("#ff9a9a","#5f5fa4","#aaed94"),border="black")
legend("topright",c("24","25","26"), cex =1, fill=c("#ff9a9a","#5f5fa4","#aaed94"),title="Age")
```

line	Explanation
1	The variable hire_2007 is assign to select and store year of status which filter the year 2007 and hire year which filter the year 2007 using the table function while selecitng the age.
2	To generate a pie chart
3	To generate legend to show information of the pie chart

Data Visualization



Observation:

Based on the data visualization above, we could observe that during year 2007, the aged of hired employees are between 24 to 26 years old where most employees are hired at the age of 25.

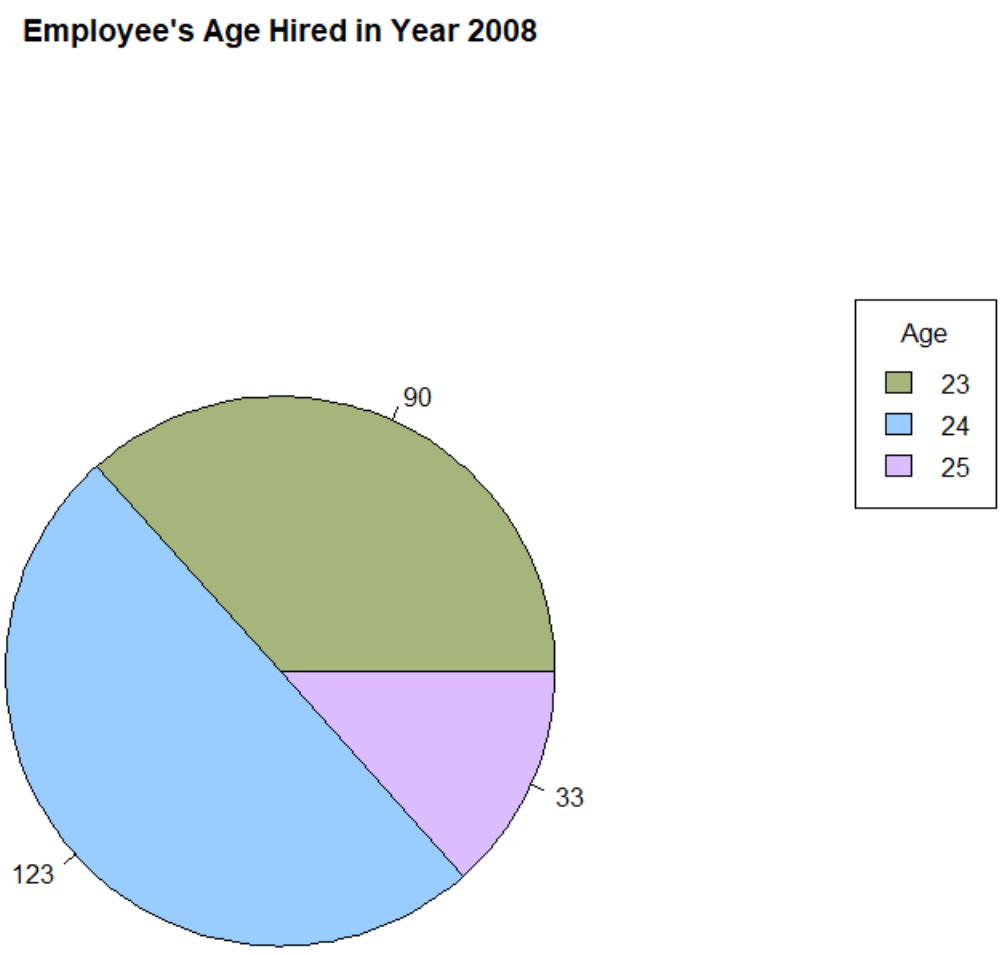
Analysis 6.3: Observation on the company's hiring interest throughout the year based on employees age in 2008

Source code

```
#in 2008
hire_2008 = table(data2 %>% filter(Year_of_Status=="2008",Hire_Year=="2008") %>% select(Age))
pie(hire_2008,label=hire_2008,main="Employee's Age Hired in Year 2008",col=c("#a6b57c","#99ccff","#dabcff"),border="black")
legend("topright",c("23","24","25"), cex =1, fill=c("#a6b57c","#99ccff","#dabcff"),title="Age")
```

line	Explanation
1	The variable hire_2008 is assign to select and store year of status which filter the year 2008 and hire year which filter the year 2008 using the table function while selecitng the age.
2	To generate a pie chart
3	To generate legend to show information of the pie chart

Data Visualization



Observation:

Based on the data visualization above, we could observe that during year 2008, the aged of hired employees are between 23 to 25 years old where most employees are hired at the age of 24.

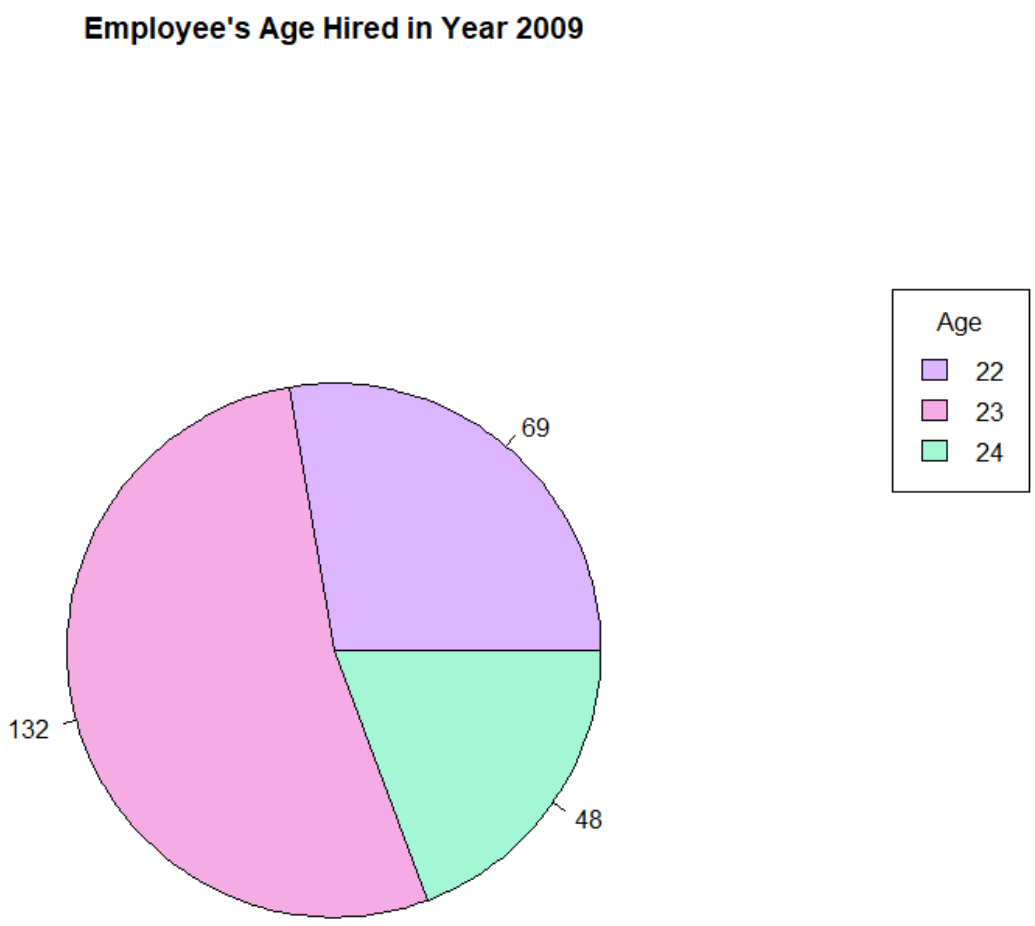
Analysis 6.4: Observation on the company's hiring interest throughout the year based on employees age in 2009

Source code

```
#in 2009
hire_2009 = table(data2 %>% filter(Year_of_Status=="2009",Hire_Year=="2009") %>% select(Age))
pie(hire_2009,label=hire_2009,main="Employee's Age Hired in Year 2009",col=c("#dbb5fd","#f5abe4","#a4f7d4"),border="black")
legend("topright",c("22","23","24"), cex =1, fill=c("#dbb5fd","#f5abe4","#a4f7d4"),title="Age")
```

line	Explanation
1	The variable hire_2009 is assign to select and store year of status which filter the year 2009 and hire year which filter the year 2009 using the table function while selecitng the age.
2	To generate a pie chart
3	To generate legend to show information of the pie chart

Data Visualization



Observation:

Based on the data visualization above, we could observe that during year 2009, the aged of hired employees are between 22 to 24 years old where most employees are hired at the age of 23.

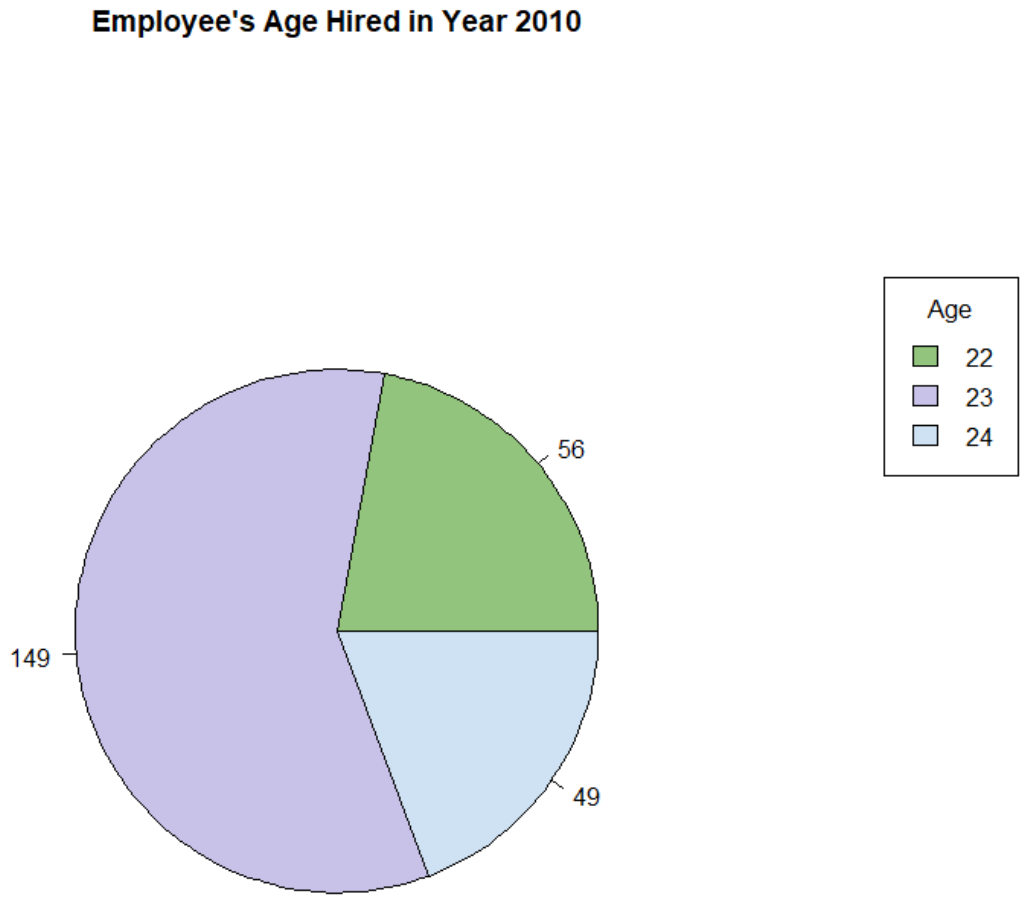
Analysis 6.5: Observation on the company's hiring interest throughout the year based on employees age in 2010

Source code

```
#in 2010
hire_2010 = table(data2 %>% filter(Year_of_Status=="2010",Hire_Year=="2010") %>% select(Age))
pie(hire_2010,label=hire_2010,main="Employee's Age Hired in Year 2010",col=c("#93c47d","#c9c2e8","#cfe2f3"),border="black")
legend("topright",c("22","23","24"), cex =1, fill=c("#93c47d","#c9c2e8","#cfe2f3"),title="Age")
```

line	Explanation
1	The variable hire_2010 is assign to select and store year of status which filter the year 2010 and hire year which filter the year 2010 using the table function while selecitng the age.
2	To generate a pie chart
3	To generate legend to show information of the pie chart

Data Visualization



Observation:

Based on the data visualization above, we could observe that during year 2010, the aged of hired employees are between 22 to 24 years old where most employees are hired at the age of 23.

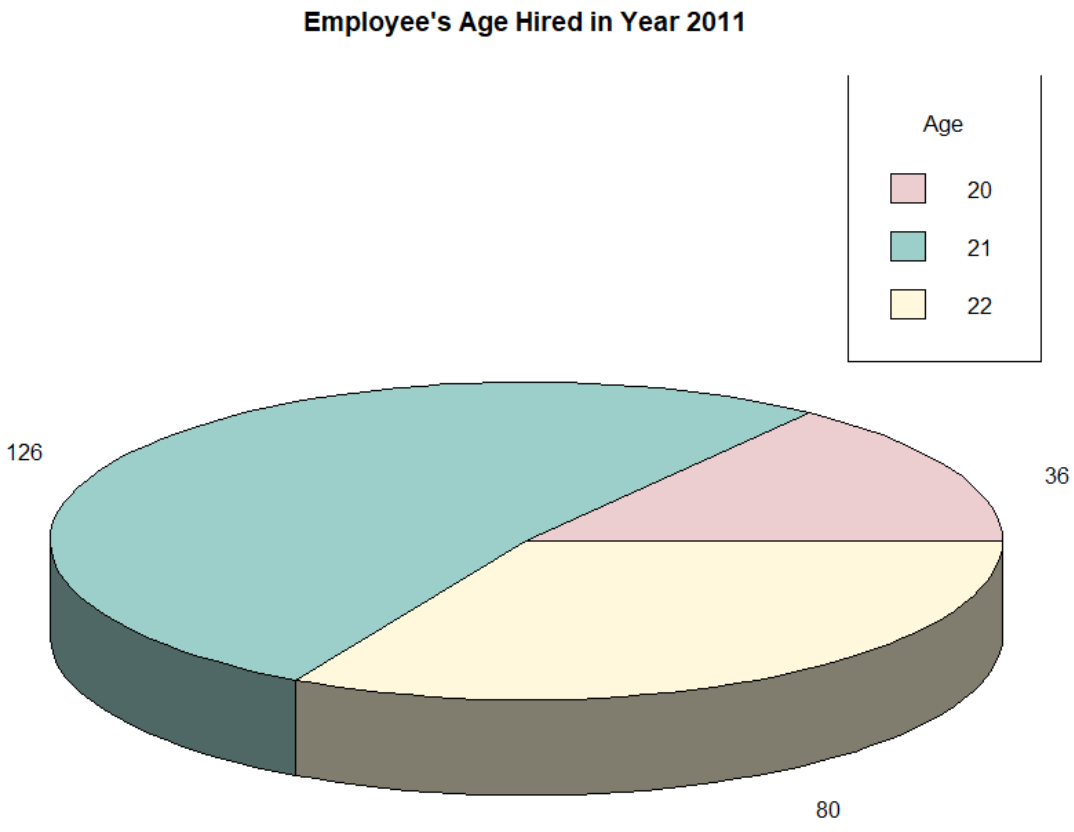
Analysis 6.6: Observation on the company's hiring interest throughout the year based on employees age in 2011

Source code

```
#in 2011
hire_2011 = table(data2 %>% filter(Year_of_Status=="2011",Hire_Year=="2011") %>% select(Age))
pie3D(hire_2011,col=c("#eccdd0","#9dcfca","#fff8dc"),main="Employee's Age Hired in Year 2011",
      border="black",shade=0.5,labels=hire_2011,labelcex = 1)
legend("topright",c("20","21","22"), cex = 1, fill=c("#eccdd0","#9dcfca","#fff8dc"),title="Age")
```

line	Explanation
1	The variable hire_2011 is assign to select and store year of status which filter the year 2011 and hire year which filter the year 2011 using the table function while selecitng the age.
2	To generate a 3D pie chart
3	To generate legend to show information of the pie chart

Data Visualization



Observation:

Based on the data visualization above, we could observe that during year 2011, the aged of hired employees are between 20 to 22 years old where most employees are hired at the age of 21.

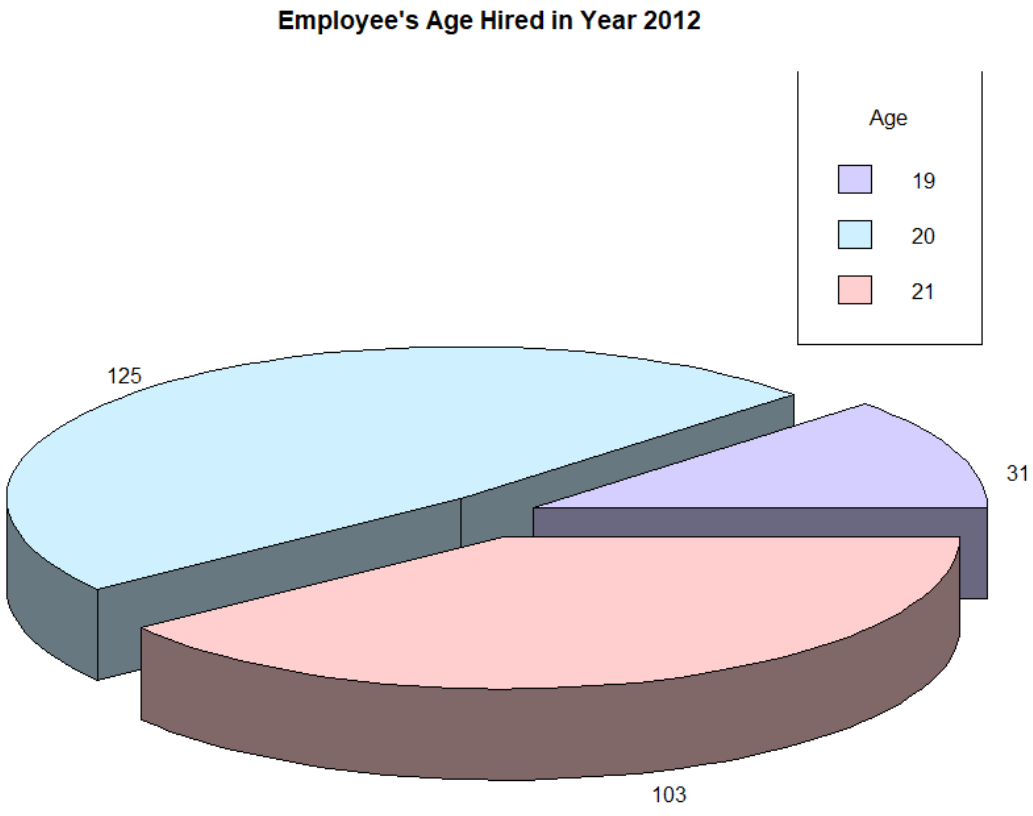
Analysis 6.7: Observation on the company's hiring interest throughout the year based on employees age in 2012

Source code

```
#in 2012
hire_2012 = table(data2 %>% filter(Year_of_Status=="2012",Hire_Year=="2012") %>% select(Age))
pie3D(hire_2012,col=c("#d4cfff","#cff0ff","#ffcfcf"),main="Employee's Age Hired in Year 2012",
      border="black",shade=0.5,labels=hire_2012,labelcex = 1,explode = 0.1)
legend("topright",c("19","20","21"), cex = 1, fill=c("#d4cfff","#cff0ff","#ffcfcf"),title="Age")
```

line	Explanation
1	The variable hire_2012 is assign to select and store year of status which filter the year 2012 and hire year which filter the year 2012 using the table function while selecitng the age.
2	To generate a 3D pie chart
3	To generate legend to show information of the pie chart

Data Visualization



Observation:

Based on the data visualization above, we could observe that during year 2012, the aged of hired employees are between 19 to 21 years old where most employees are hired at the age of 20.

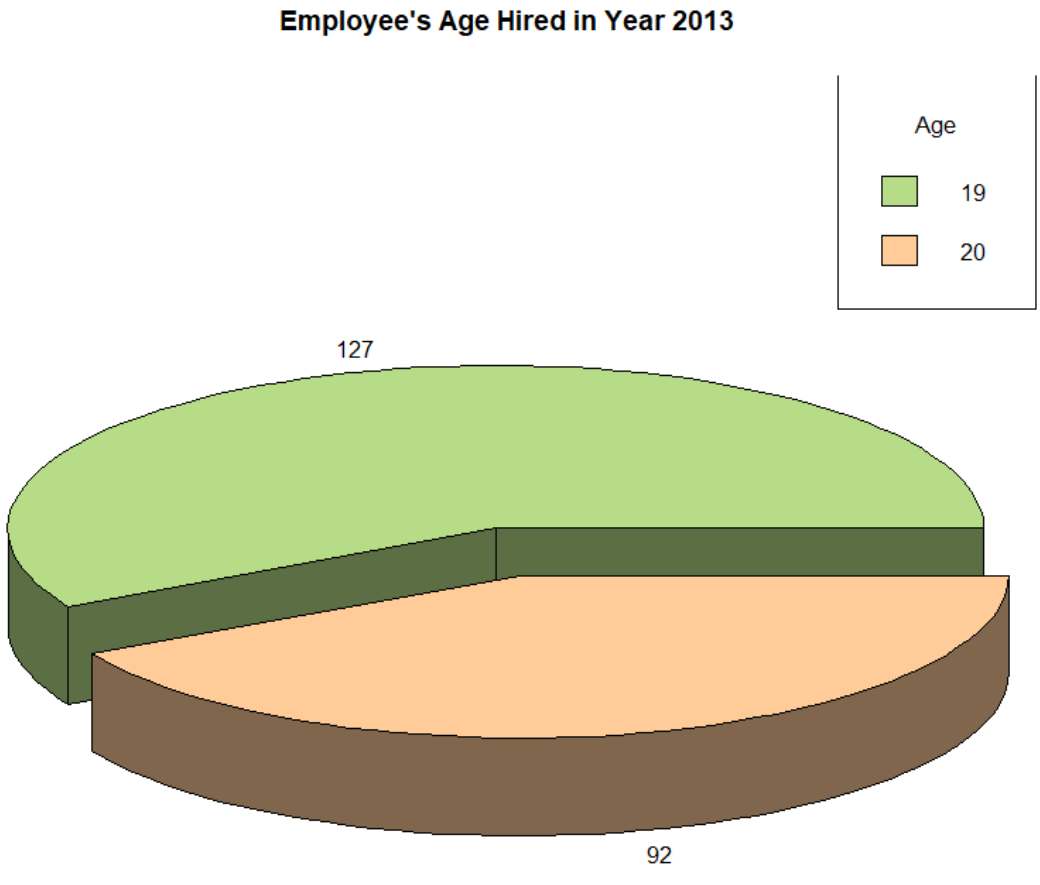
Analysis 6.8: Observation on the company's hiring interest throughout the year based on empoyees age in 2013

Source code

```
#in 2013
hire_2013 = table(data2 %>% filter(Year_of_Status=="2013",Hire_Year=="2013") %>% select(Age))
pie3D(hire_2013,col=c("#b7dc87","#ffcc99"),main="Employee's Age Hired in Year 2013",
      border="black",shade=0.5,labels=hire_2013,labelcex = 1,explode = 0.1)
legend("topright",c("19","20"), cex = 1, fill=c("#b7dc87","#ffcc99"),title="Age")
```

line	Explanation
1	The variable hire_2013 is assign to select and store year of status which filter the year 2013 and hire year which filter the year 2013 using the table function while selecitng the age.
2	To generate a 3D pie chart
3	To generate legend to show information of the pie chart

Data Visualization



Observation:

Based on the data visualization above, we could observe that during year 2013, the aged of hired employees are between 19 to 20 years old where most employees are hired at the age of 19.

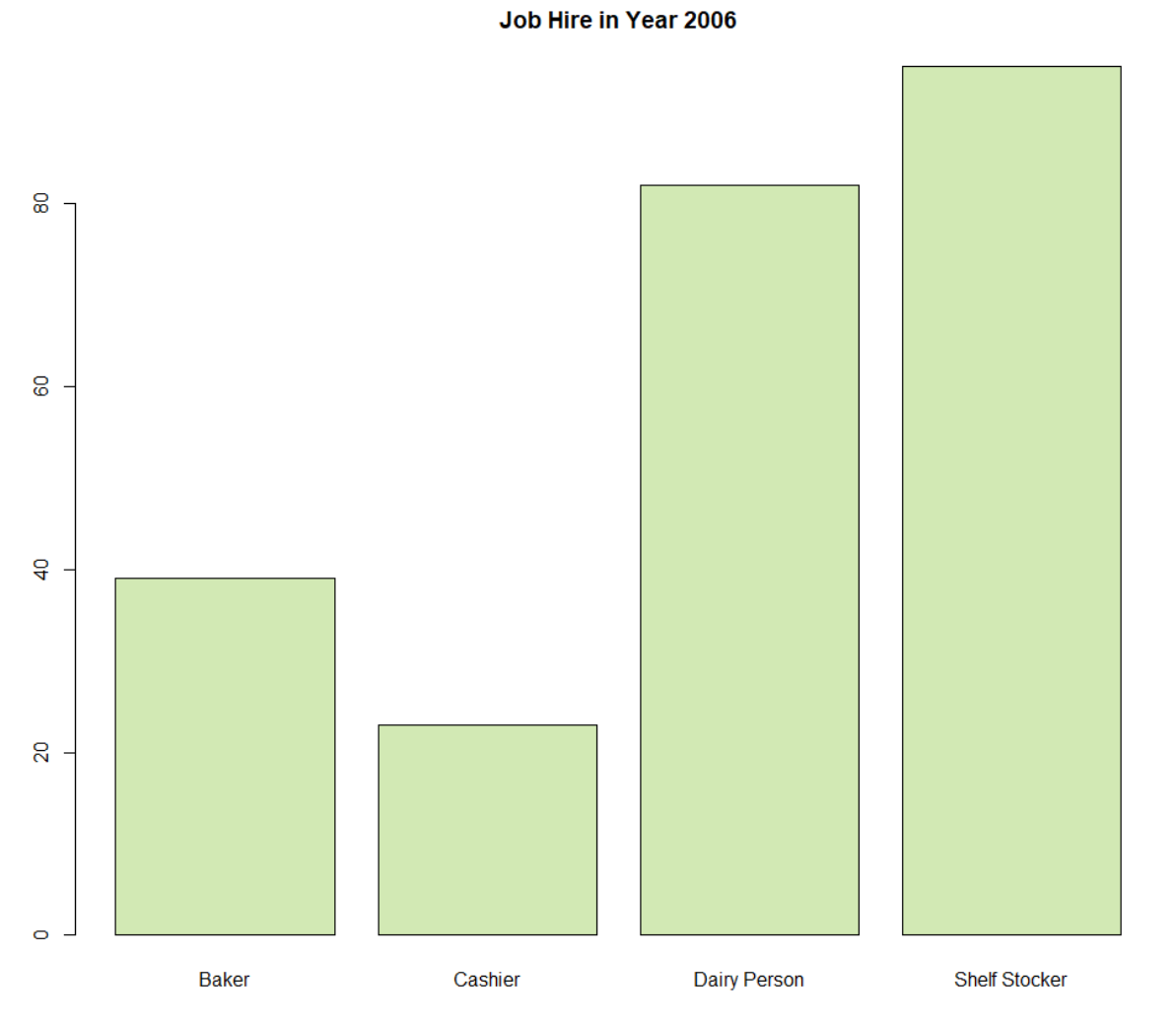
Analysis 6.9: Observation on the company's hiring interest throughout the year based on employees job in 2006

Source code

```
#job
#2006
job_2006 = data2 %>% filter(Year_of_Status=="2006",Hire_Year=="2006") %>% group_by(Job) %>%summarise(Emplyee=n())
barplot(height=job_2006$Emplyee, beside = TRUE,main = "Job Hire in Year 2006",names.arg = c("Baker","Cashier","Dairy Person","Shelf Stocker"),
        col = "#d2e9b4")
```

line	Explanation
1	To assign the variable job_2006 with summarize data year of status and hiring year filter to only year 2006 while grouping them by job
2	To generate the bar chart

Data Visualization



Observation:

Based on the data visualization above, we could observe that during the year 2006, the company hired shelf stocker the most. Other than that, the company also hired decent amount of bakers, cashier and dairy person.

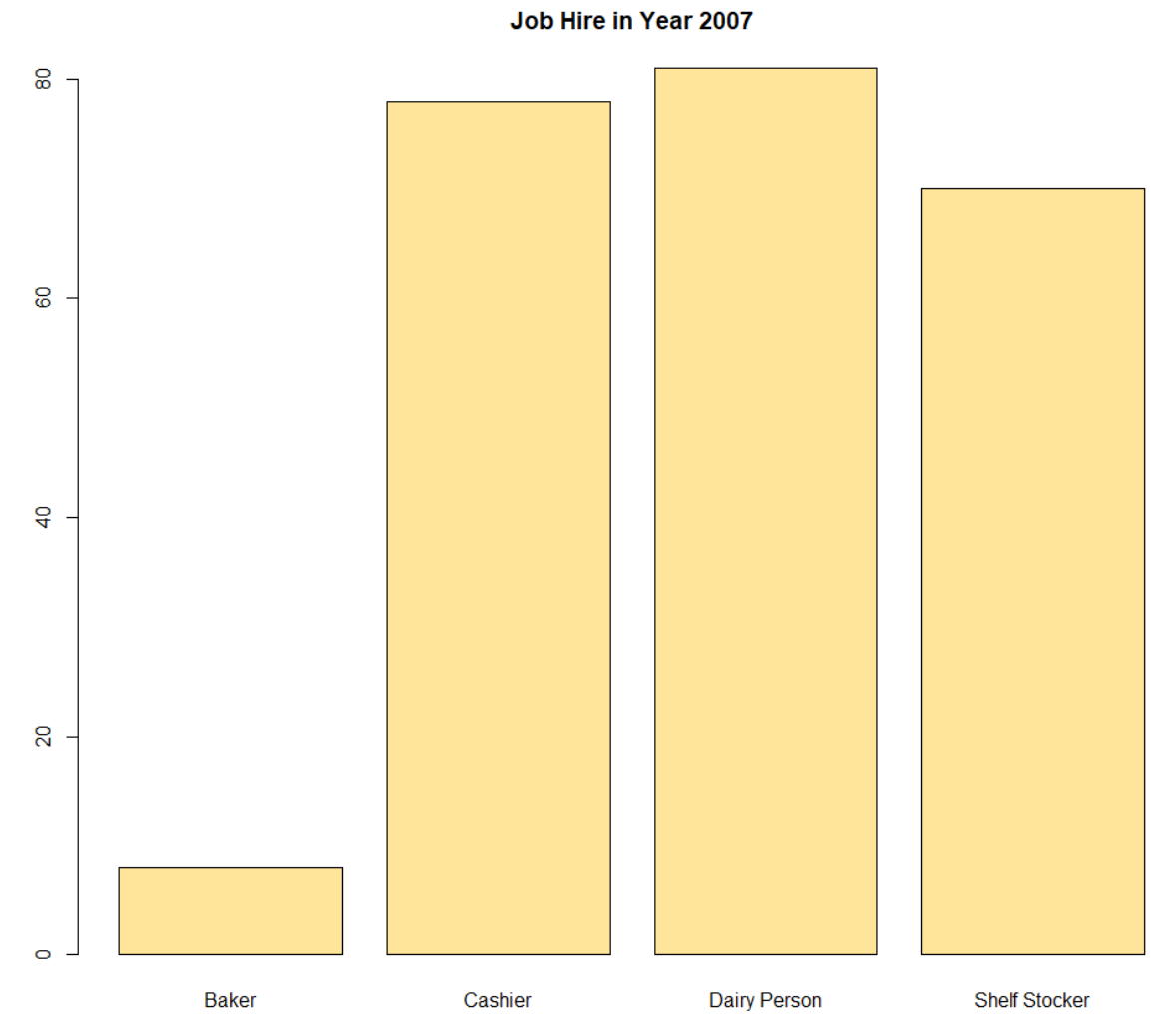
Analysis 6.10: Observation on the company's hiring interest throughout the year based on employees job in 2007

Source code

```
#2007
job_2007 = data2 %>% filter(Year_of_Status=="2007",Hire_Year=="2007") %>% group_by(Job) %>%summarise(Emplyee=n())
barplot(height=job_2007$Emplyee, beside = TRUE,main = "Job Hire in Year 2007",names.arg = job_2007$Job,
        col = "#ffe599")
```

line	Explanation
1	To assign the variable job_2007 with summarize data year of status and hiring year filter to only year 2007 while grouping them by job
2	To generate the bar chart

Data Visualization



Observation:

Based on the data visualization above, we could observe that during the year 2007, the company hired dairy person the most. Other than that, the company also hired decent amount of bakers, cashier and shelf stockers.

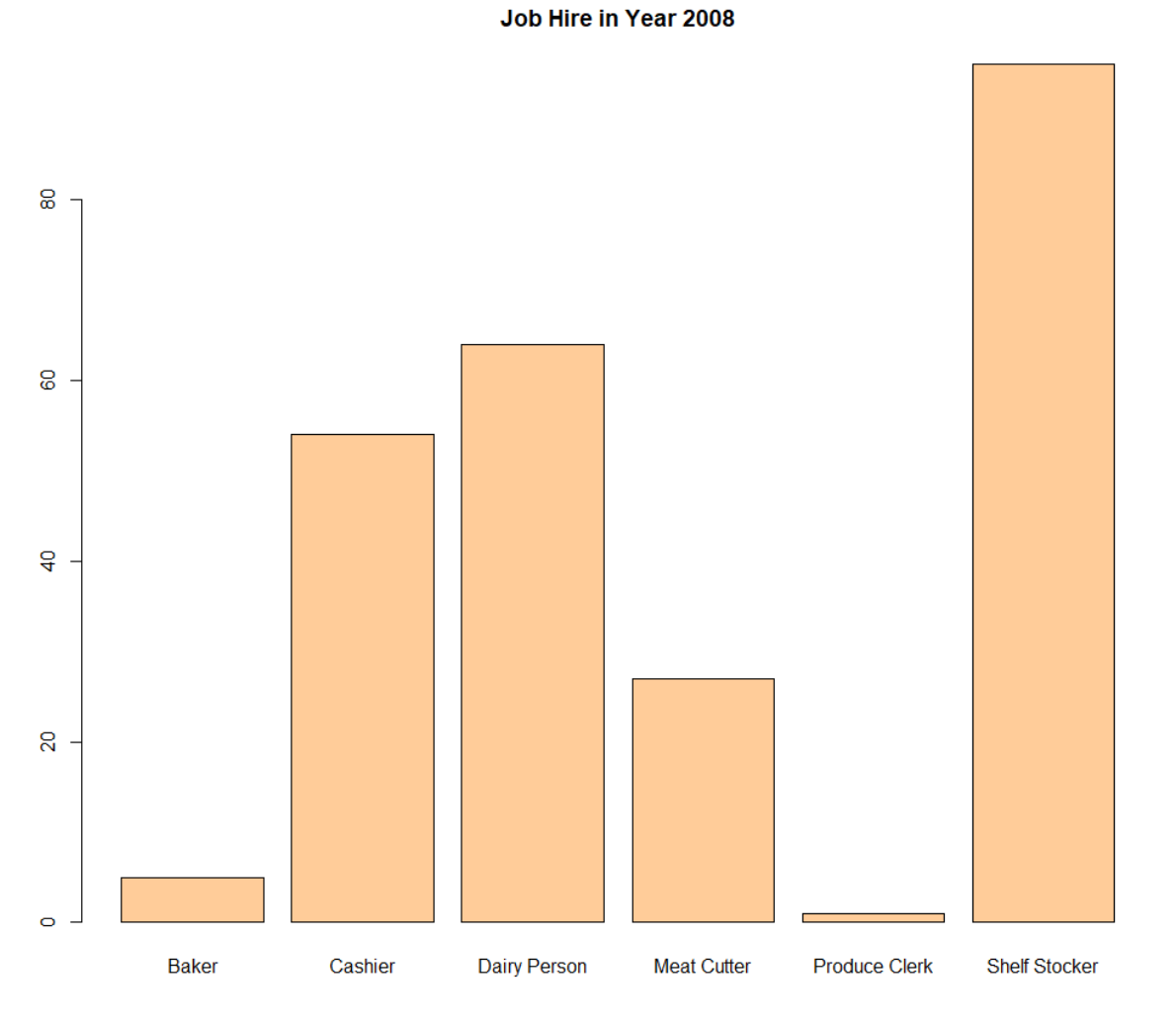
Analysis 6.11: Observation on the company's hiring interest throughout the year based on employees job in 2008

Source code

```
#2008
job_2008 = data2 %>% filter(Year_of_Status=="2008",Hire_Year=="2008") %>% group_by(Job) %>%summarise(Emplyee=n())
barplot(height=job_2008$Emplyee, beside = TRUE,main = "Job Hire in Year 2008",names.arg = job_2008$Job,
        col = "#ffcc99")
```

line	Explanation
1	To assign the variable job_2008 with summarize data year of status and hiring year filter to only year 2008 while grouping them by job
2	To generate the bar chart

Data Visualization



Observation:

Based on the data visualization above, we could observe that during the year 2008, the company hired shelf stocker the most. Other than that, the company also hired decent amount of bakers, cashier, dairy person , meat cutters and produce clerk.

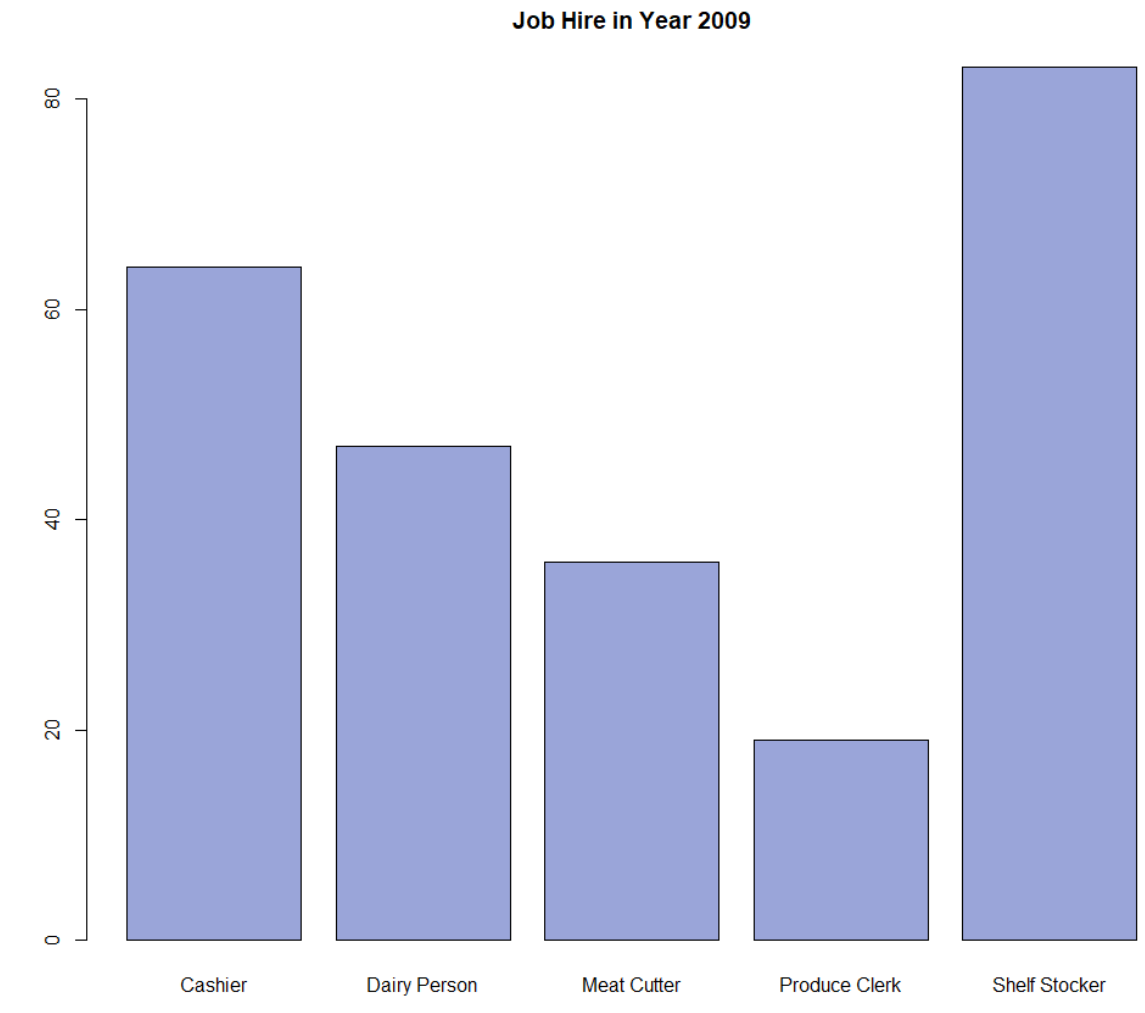
Analysis 6.12: Observation on the company's hiring interest throughout the year based on employees job in 2009

Source code

```
#2009
job_2009 = data2 %>% filter(Year_of_Status=="2009",Hire_Year=="2009") %>% group_by(Job) %>%summarise(Emplyee=n())
barplot(height=job_2009$Emplyee, beside = TRUE,main = "Job Hire in Year 2009",names.arg = job_2009$Job,
        col = "#9aa5d9")
```

line	Explanation
1	To assign the variable job_2009 with summarize data year of status and hiring year filter to only year 2009 while grouping them by job
2	To generate the bar chart

Data Visualization



Observation:

Based on the data visualization above, we could observe that during the year 2009, the company hired shelf stocker the most. Other than that, the company also hired decent amount of bakers, cashier, dairy person , meat cutters and produce clerk.

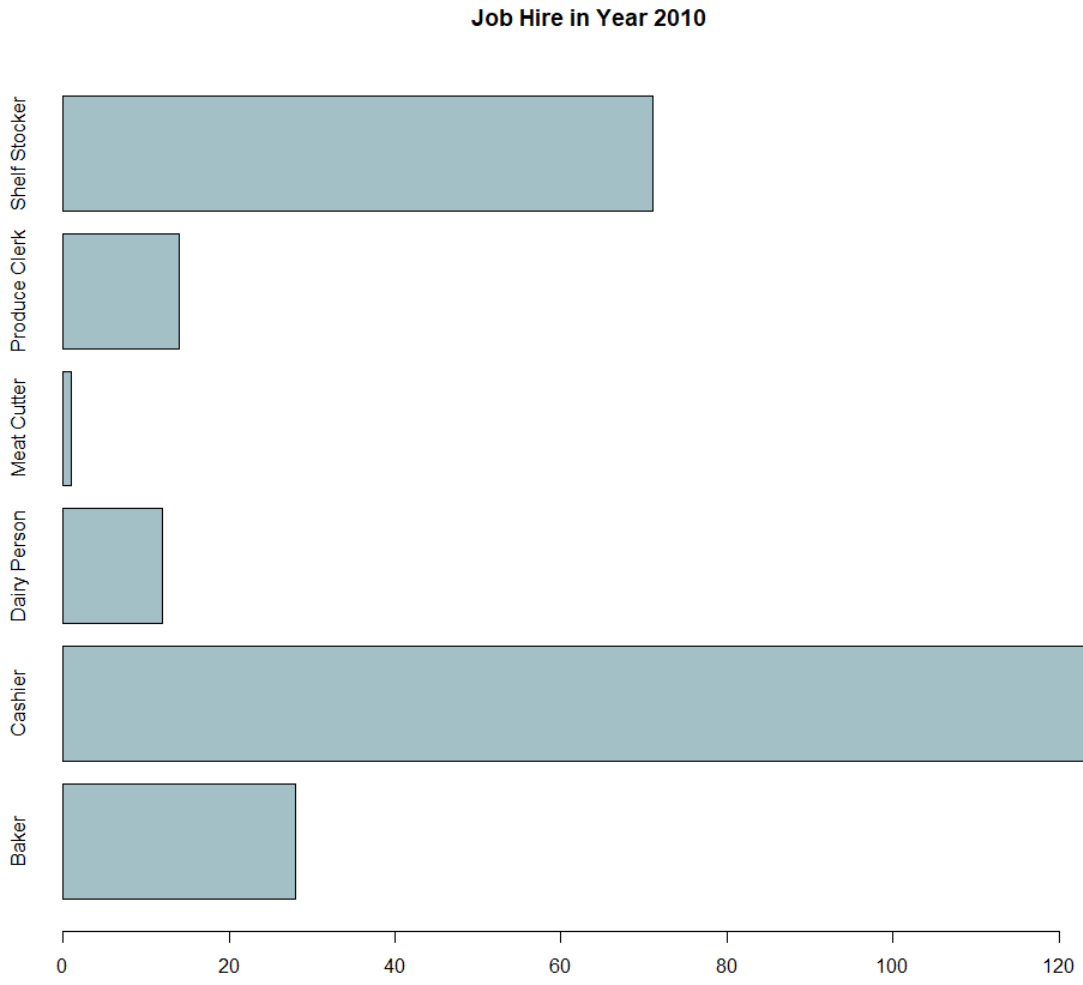
Analysis 6.13: Observation on the company's hiring interest throughout the year based on employees job in 2010

Source code

```
#2010
job_2010 = data2 %>% filter(Year_of_Status=="2010",Hire_Year=="2010") %>% group_by(Job) %>%summarise(Emplyee=n())
barplot(height=job_2010$Emplyee, beside = TRUE,main = "Job Hire in Year 2010",names.arg = job_2010$Job,
        col = "#a3c0c6",horiz = TRUE)
```

line	Explanation
1	To assign the variable job_2010 with summarize data year of status and hiring year filter to only year 2010 while grouping them by job
2	To generate the bar chart

Data Visualization



Observation:

Based on the data visualization above, we could observe that during the year 2010, the company hired cashiers the most. Other than that, the company also hired decent amount of shelf stockers, bakers, dairy person , meat cutters and produce clerk.

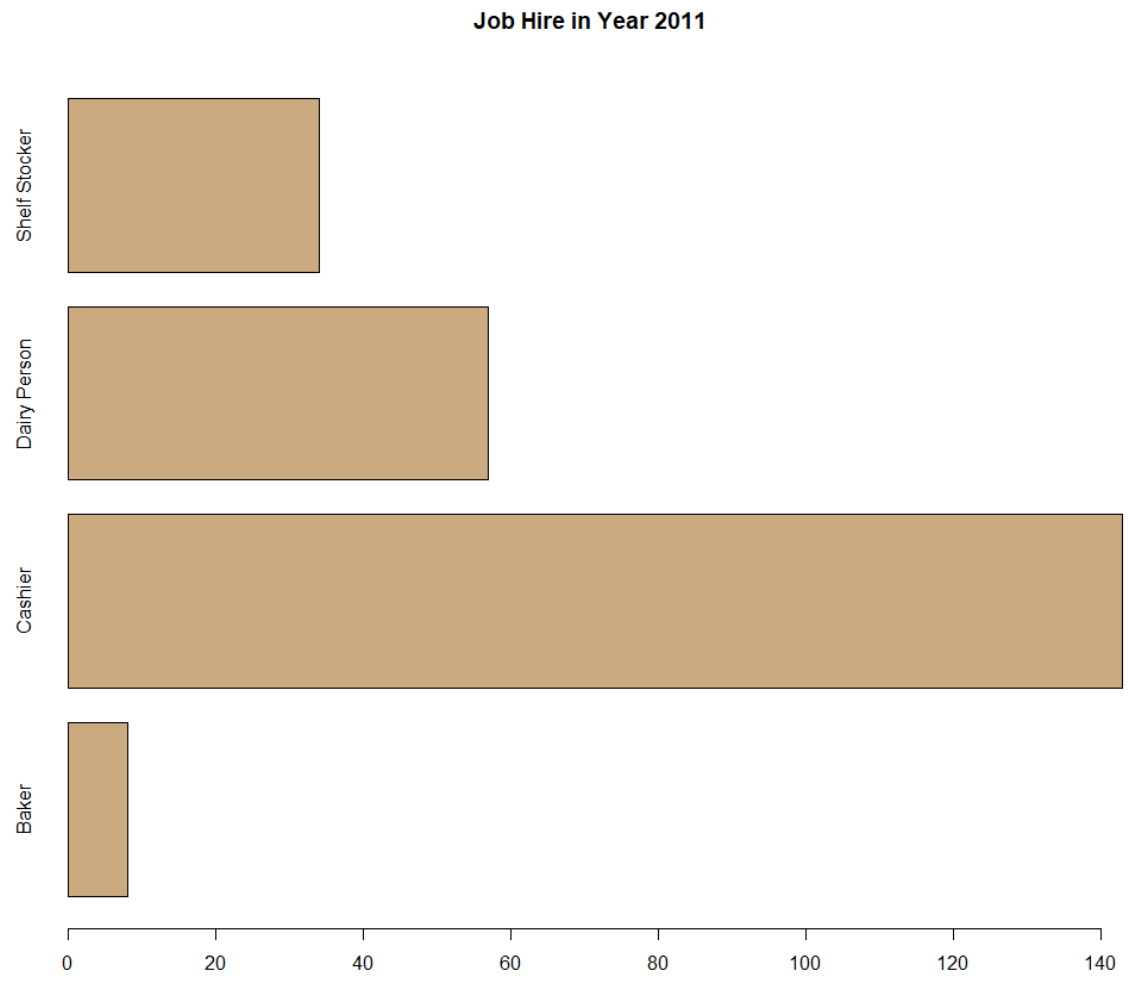
Analysis 6.14: Observation on the company's hiring interest throughout the year based on employees job in 2011

Source code

```
#2011
job_2011 = data2 %>% filter(Year_of_Status=="2011",Hire_Year=="2011") %>% group_by(Job) %>%summarise(Emplyee=n())
barplot(height=job_2011$Emplyee, beside = TRUE,main = "Job Hire in Year 2011",names.arg = job_2011$Job,
        col = "#ccaa80",horiz = TRUE)
```

line	Explanation
1	To assign the variable job_2011 with summarize data year of status and hiring year filter to only year 2011 while grouping them by job
2	To generate the bar chart

Data Visualization



Observation:

Based on the data visualization above, we could observe that during the year 2011, the company hired cashiers the most. Other than that, the company also hired decent amount of shelf stockers, bakers and dairy person.

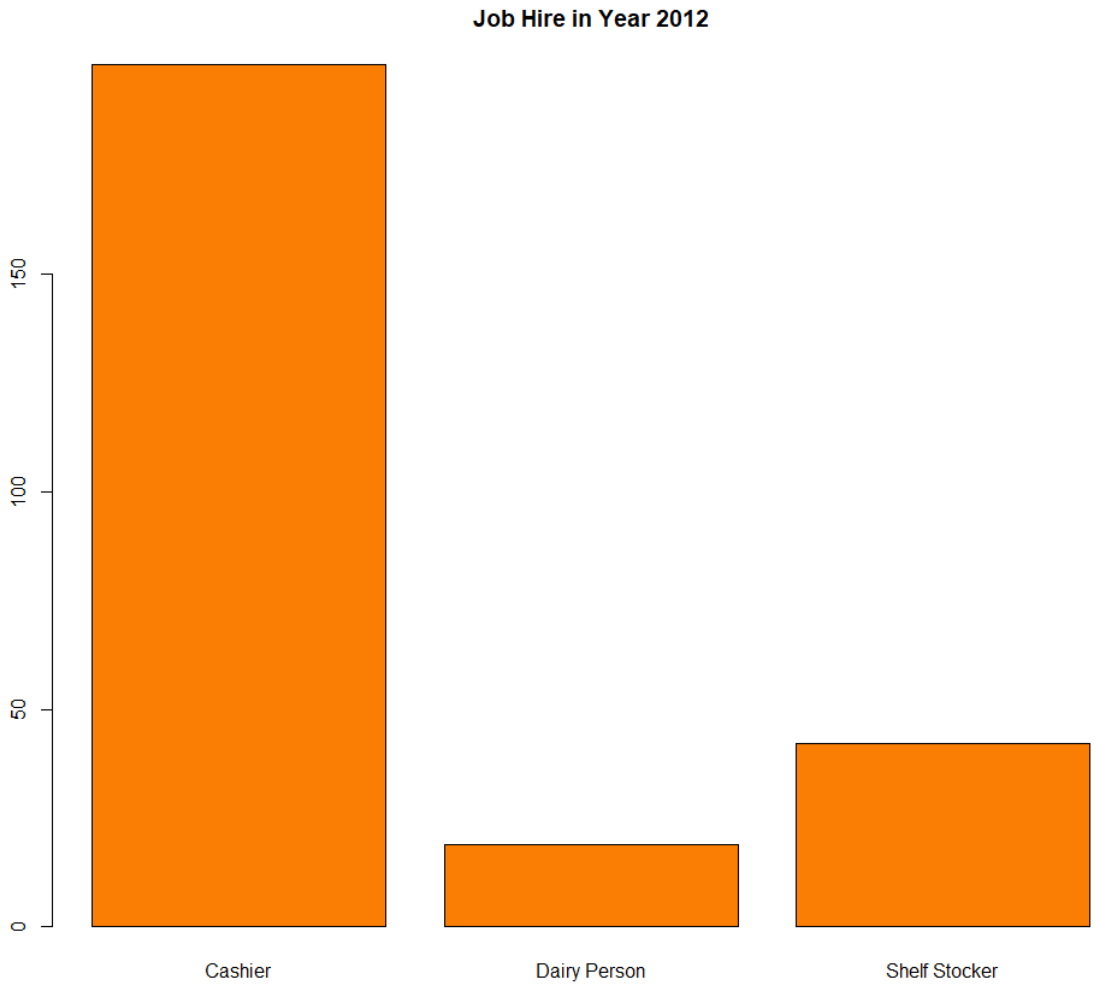
Analysis 6.15: Observation on the company's hiring interest throughout the year based on employees job in 2012

Source code

```
#2012
job_2012 = data2 %>% filter(Year_of_Status=="2012",Hire_Year=="2012") %>% group_by(Job) %>%summarise(Emplyee=n())
barplot(height=job_2012$Emplyee, beside = TRUE,main = "Job Hire in Year 2012",names.arg = job_2012$Job,
        col = "#fa7e04")
```

line	Explanation
1	To assign the variable job_2012 with summarize data year of status and hiring year filter to only year 2012 while grouping them by job
2	To generate the bar chart

Data Visualization



Observation:

Based on the data visualization above, we could observe that during the year 2012, the company hired cashiers the most. Other than that, the company also hired decent amount of shelf stockers and dairy persion.

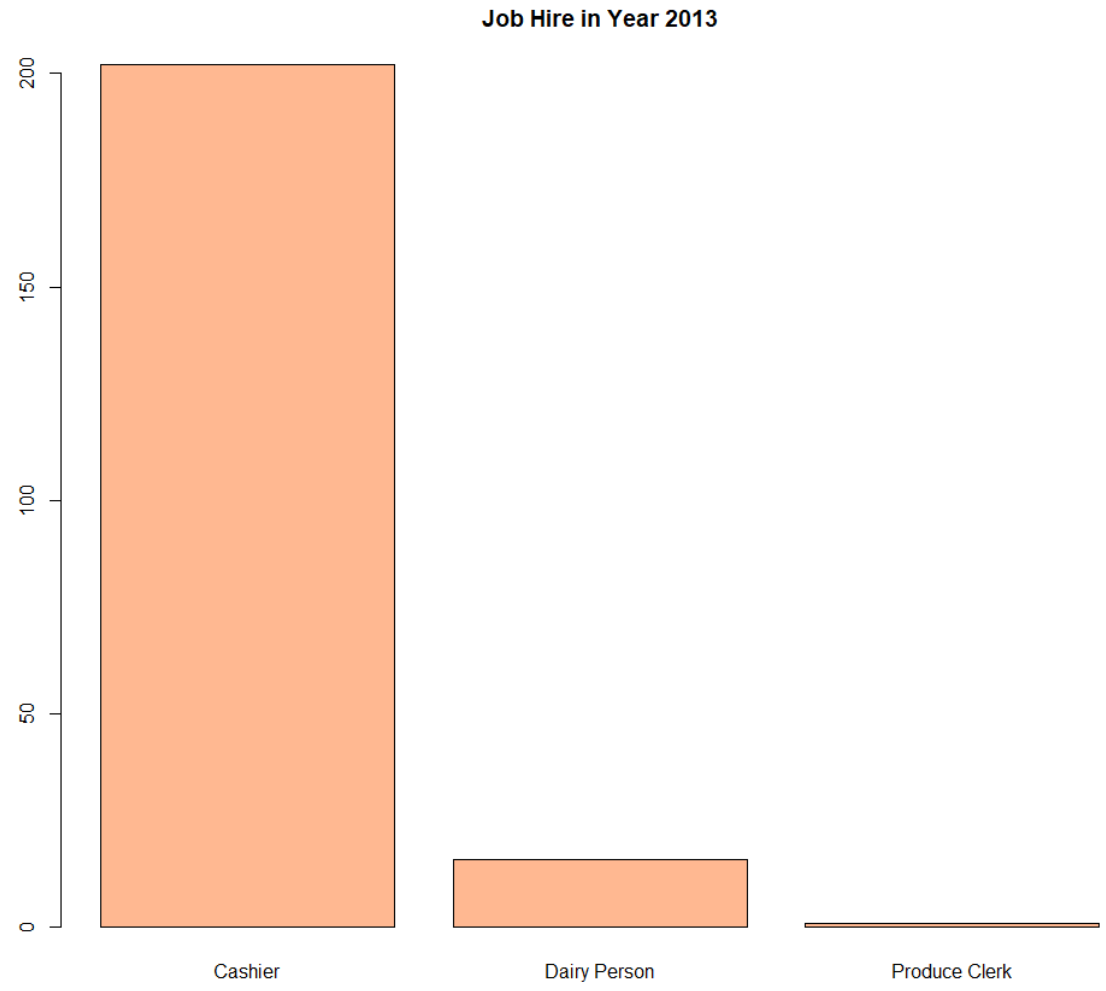
Analysis 6.16: Observation on the company's hiring interest throughout the year based on employees job in 2013

Source code

```
#2013
job_2013 = data2 %>% filter(Year_of_Status=="2013",Hire_Year=="2013") %>% group_by(Job) %>%summarise(Emplyee=n())
barplot(height=job_2013$Emplyee, beside = TRUE,main = "Job Hire in Year 2013",names.arg = job_2013$Job,
        col = "#ffb891")
```

line	Explanation
1	To assign the variable job_2013 with summarize data year of status and hiring year filter to only year 2013 while grouping them by job
2	To generate the bar chart

Data Visualization



Observation:

Based on the data visualization above, we could observe that during the year 2013, the company hired cashiers the most. Other than that, the company also hired decent amount of producce clerk and dairy person.

Conclusion of Analysis 6: Observation on the company's hiring interest throughout the year

After conducting various analysis based on: -

- Observation on the company's hiring interest throughout the year based on employees age in 2006
- Observation on the company's hiring interest throughout the year based on employees age in 2007
- Observation on the company's hiring interest throughout the year based on employees age in 2008
- Observation on the company's hiring interest throughout the year based on employees age in 2009
- Observation on the company's hiring interest throughout the year based on employees age in 2010
- Observation on the company's hiring interest throughout the year based on employees age in 2011
- Observation on the company's hiring interest throughout the year based on employees age in 2012
- Observation on the company's hiring interest throughout the year based on employees age in 2013
- Observation on the company's hiring interest throughout the year based on employees' job in 2006
- Observation on the company's hiring interest throughout the year based on employees' job in 2007
- Observation on the company's hiring interest throughout the year based on employees' job in 2008
- Observation on the company's hiring interest throughout the year based on employees' job in 2009
- Observation on the company's hiring interest throughout the year based on employees' job in 2010
- Observation on the company's hiring interest throughout the year based on employees' job in 2011
- Observation on the company's hiring interest throughout the year based on employees' job in 2012

- Observation on the company's hiring interest throughout the year based on employees' job in 2013

It could be concluded that: -

- From the year 2006 to year 2013, most emloyees hired by the company are mostly between the age 19 to 27 years old
- From the year 2006 to year 2013, most of the hiring jobs are shelf stocker, dairy person and cashier

Analysis 7: Observation on the company's internal structure

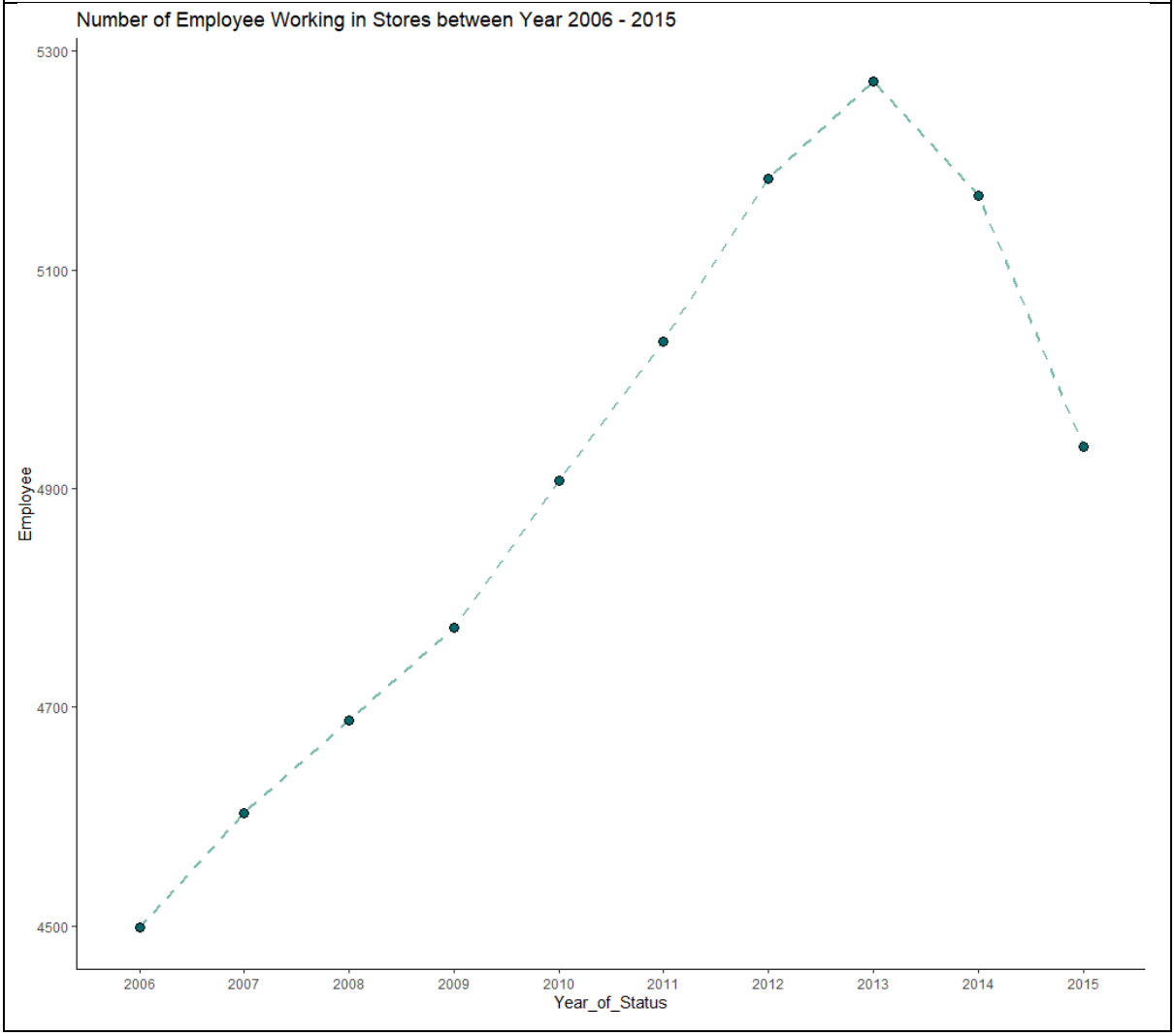
Analysis 7.1: Observation on the company's internal structure based on the company's stores

Source code

```
#7. How many stores and Headoffice?
#store
data2 %>% filter(Business_Unit=="STORES")%>%group_by(Year_of_Status,Business_Unit) %>%summarise(Employee=n()) %>% ggplot(aes(x=Year_of_Status,y=Employee,group=Business_Unit))+
  geom_line(color="#69b3a2", size=1, alpha=0.9, linetype=2)+geom_point(shape=21, color="black", fill="#00676a", size=3)+
  ggtitle("Number of Employee Working in Stores between Year 2006 - 2015")+theme_classic()
```

line	Explanation
1	To summarie the data business unit to filter only employee from store while grouping them by their year of status and business unit
2	To generate the line graph
3	To generate title for the line graph

Data Visualization



Observation:

Based on the data visualization above, we could observe that the number of employee working in stores between year 2006 to 2013 are in a steady increase with the year 2013 having the most employee(5300), then at the year 2014 to 2015, there have been a decline in the number of employee working in store.

Analysis 7.2: Observation on the company's internal structure based on the company's headoffice

Source code

```
#Headoffice
data2 %>% filter(Business_Unit=="HEADOFFICE")%>%group_by(Year_of_Status,Business_Unit) %>%summarise(Employee=n()) %>% ggplot(aes(x=Year_of_Status,y=Employee,group=Business_Unit))+
  geom_line(color="black", size=1, alpha=0.9, linetype=2)+geom_point(shape=23, color="#c71585", fill="#e485b4", size=2)+
  ggtitle("Number of Employee Working in Headoffice in Year 2006 - 2015")+theme_classic()
```

line	Explanation
1	To summarize data business unit to filter only headoffice while group them by their year of service and business unit
2	To generate the line chart
3	To generate the title for the line chart

Data Visualization

Observation:

Based on the data visualization above, we could observe that the number of employee working in headoffice from year 2006 to 2015 have been in a declining state.

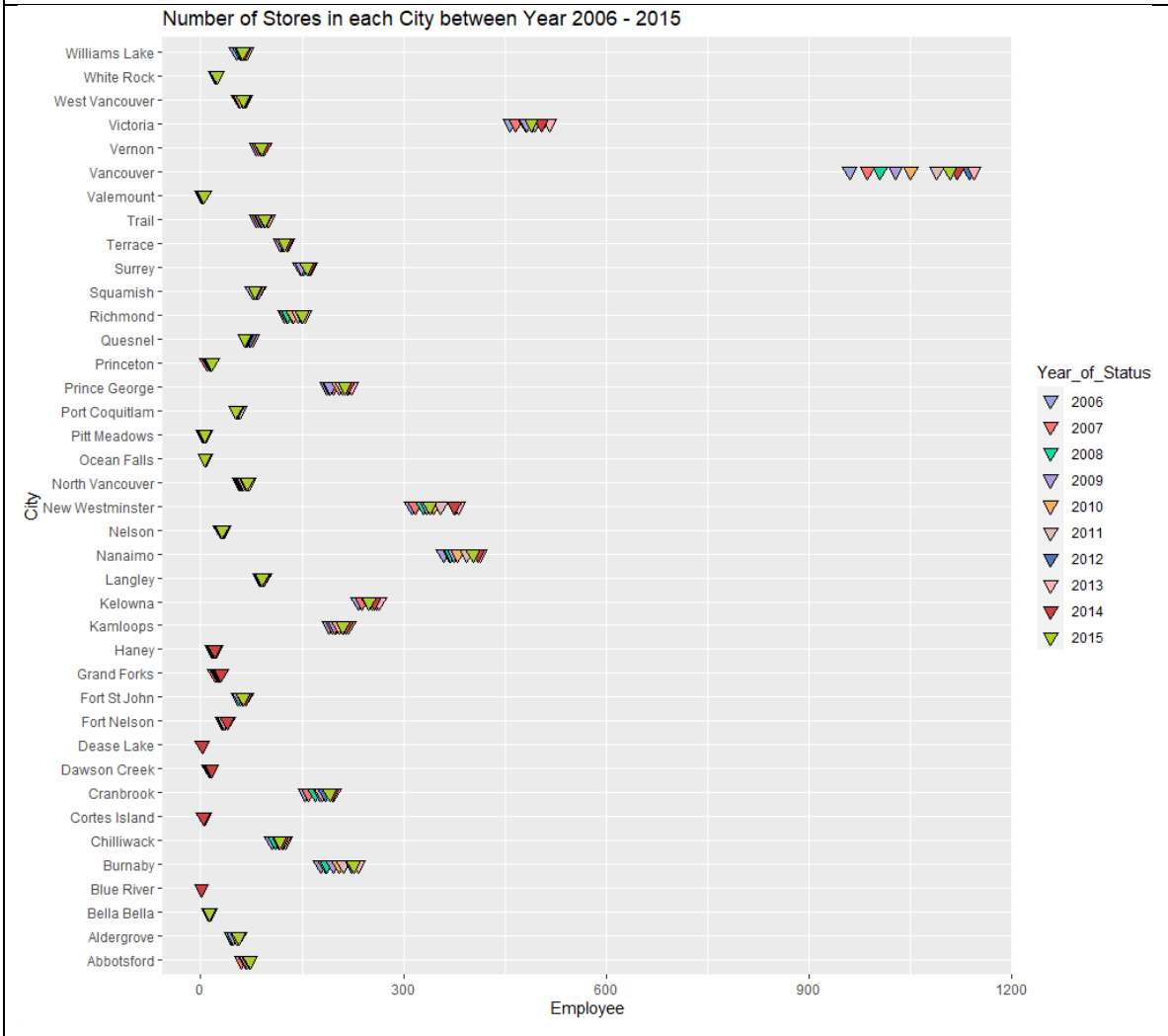
Analysis 7.3: Observation on the company's internal structure based on the company's store and city

Source code

```
#store vs City
#2006
data2 %>% filter(Business_Unit=="STORES")%>%group_by(City,Year_of_Status) %>% summarise(Employee=n()) %>%
  ggplot(aes(fill=Year_of_Status,x=City,y=Employee))+geom_point(size=3,col="black",shape=25)+
  coord_flip()+scale_fill_manual(values=c("#9aa5d9","#ff7373","#0fd99b","#ae9cde","#f7b066",
                                          "#dfbbb8","#5176be","#ffb0bb","#cc4040","#abce29","#b86e7d"))+
  ggtitle("Number of Stores in each City between Year 2006 - 2015")
```

line	Explanation
1	To summarize the data business unit to filter only stores employee while grouping thembu their city and year of status.
2-3	To generate then scatter plot graph
4	To generate the title for the scatter plot

Data Visualization



Observation:

Based on the data visualization above, we could observe that cities like vancouver have the most number of stores, cities like victoria new westminster and nanaimo also decent amount number of stores.

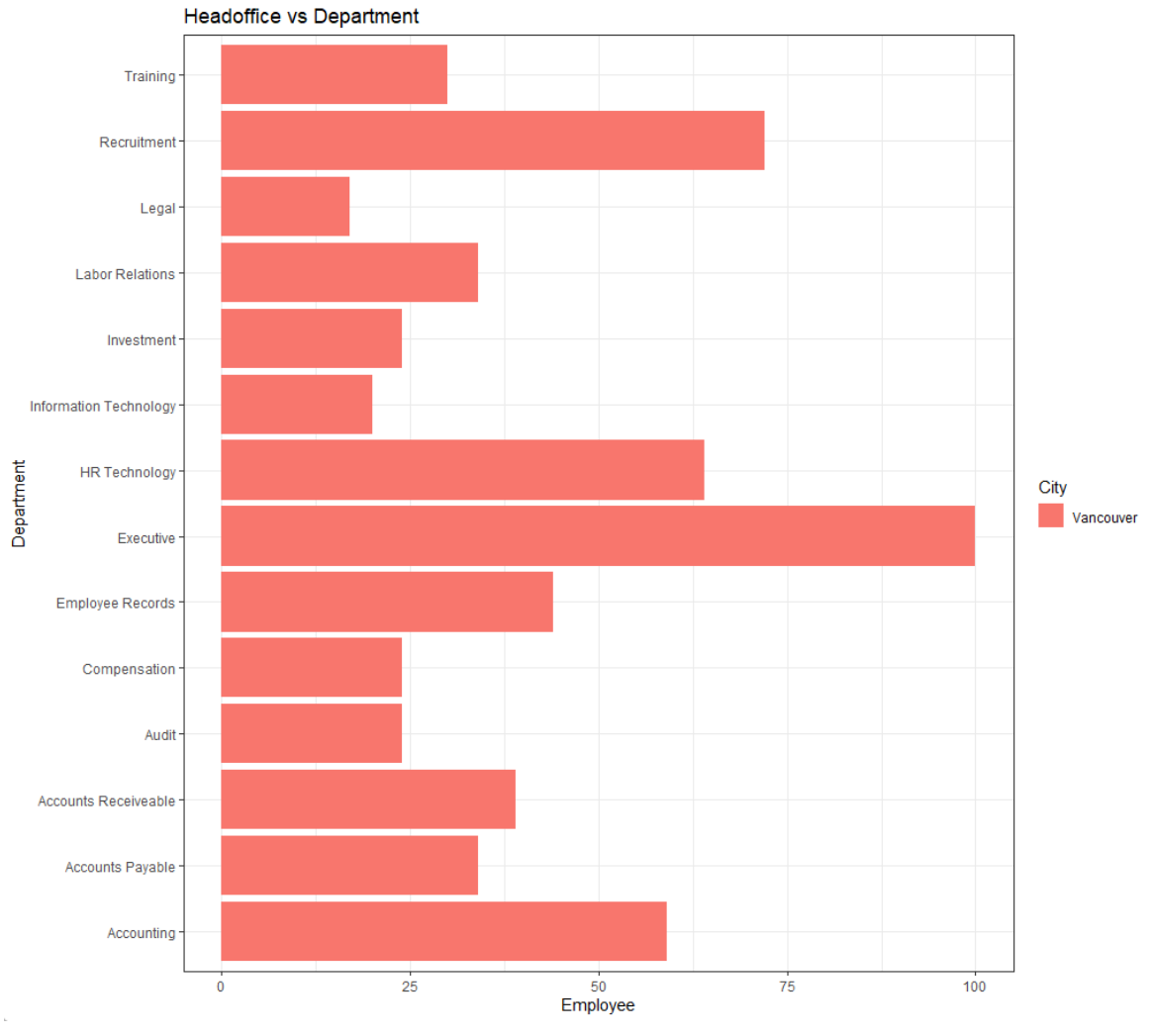
Analysis 7.4: Observation on the company's internal structure based on the company's headoffice, city and department

Source code

```
#headoffice vs city vs Department
data2 %>% filter(Business_Unit=="HEADOFFICE")%>%group_by(City,Department) %>%summarise(Employee=n()) %>%
  ggplot(aes(x=Department,y=Employee,fill=City))+geom_bar(stat="identity")+coord_flip()+
  ggtitle("Headoffice vs Department")+theme_bw()
```

line	Explanation
1	To summarize the data business unit to filter only headoffice employees while grouping them by their city and department
2	To generate the bar chart
3	To generate title for the bar chart

Data Visualization



Observation:

Based on the data visualization above, we could observe that the executive department has the most employees in the city vancouver. Moreover, recruitment, HR technology and accounting too have a moderate amount of workers.

Conclusion of Analysis 7: Observation on the voluntary termination rate of the company

After conducting various analysis based on: -

- Observation on the company's internal structure based on the company's stores
- Observation on the company's internal structure based on the company's headoffice
- Observation on the company's internal structure based on the company's store and city
- Observation on the company's internal structure based on the company's headoffice, city and department

It could be concluded that: -

- The year 2006 to 2013, the number of employees working in stores are increasing
- The number of employees working in stores are slowly decreasing at the year 2014 to 2015.
- Throughout the yer 2006 to 2015, the number of employeesin the headoffice is decreasing
- The executive department has the most employee at the headoffice placed at vancouver.

Evidence: -

It is mentions by BalanceCareer, where economic downturn are one of the major reasons for company to stop hiring and start lay off employees (BalanceCareer, 2022). Relating back to the company, the company is seemed to be overstaffed at the store during year 2013, which may result in the increase of termination of the company. Other than that, an economic downturn might be happening which result in the decrease of hiring in the headoffice departments.

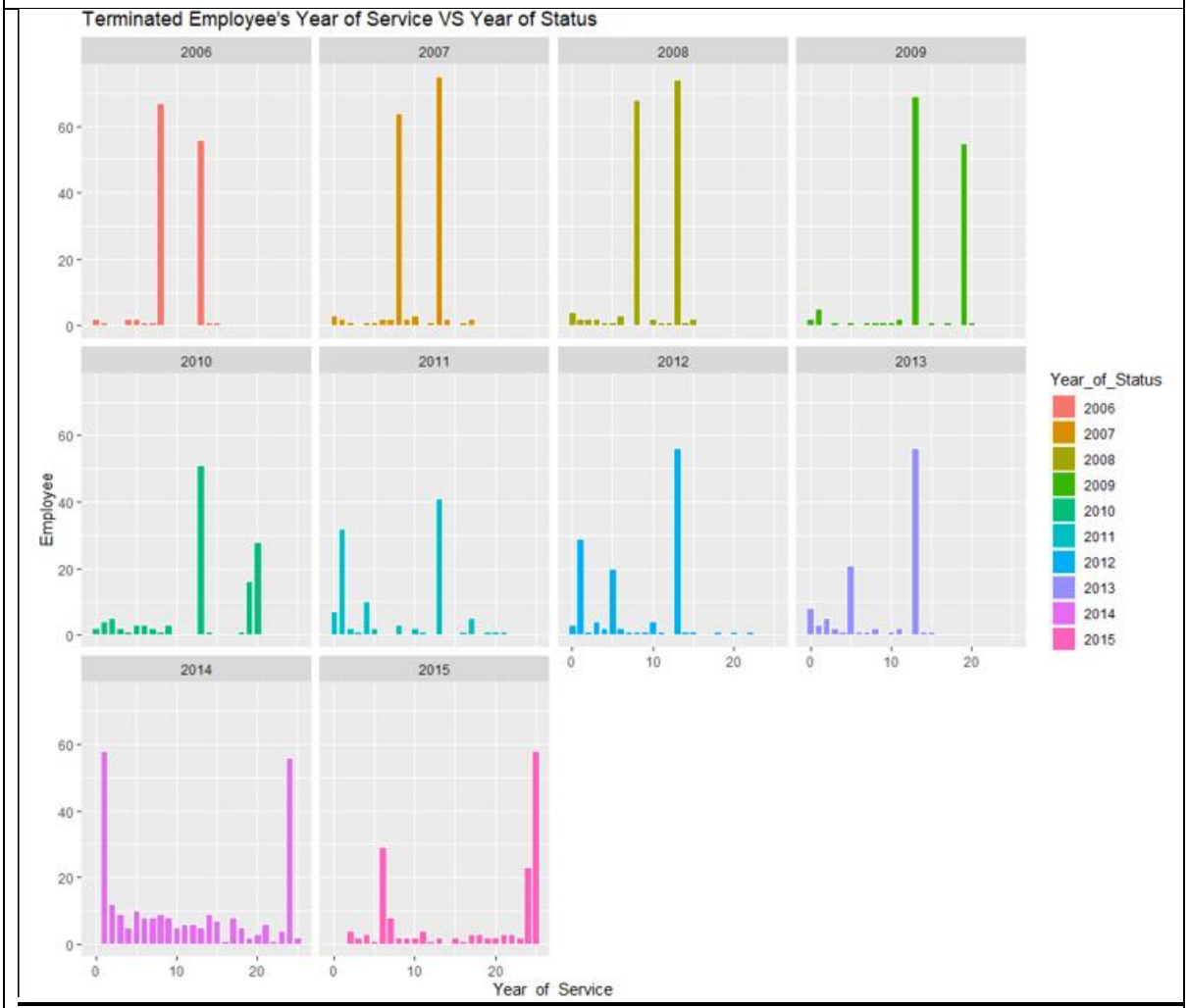
Additional Features

Facet Function

Source code

```
#length
data2 %>% filter(Status == "TERMINATED") %>% group_by(Year_of_Service,Year_of_Status) %>% summarise(Employee=n()) %>%
  ggplot(aes(x=Year_of_Service,y=Employee,fill=Year_of_Status))+geom_bar(stat="identity",col="white")+
  facet_wrap(~Year_of_Status)+
  ggtitle("Terminated Employee's Year of Service VS Year of Status")
```

Output



Explanation

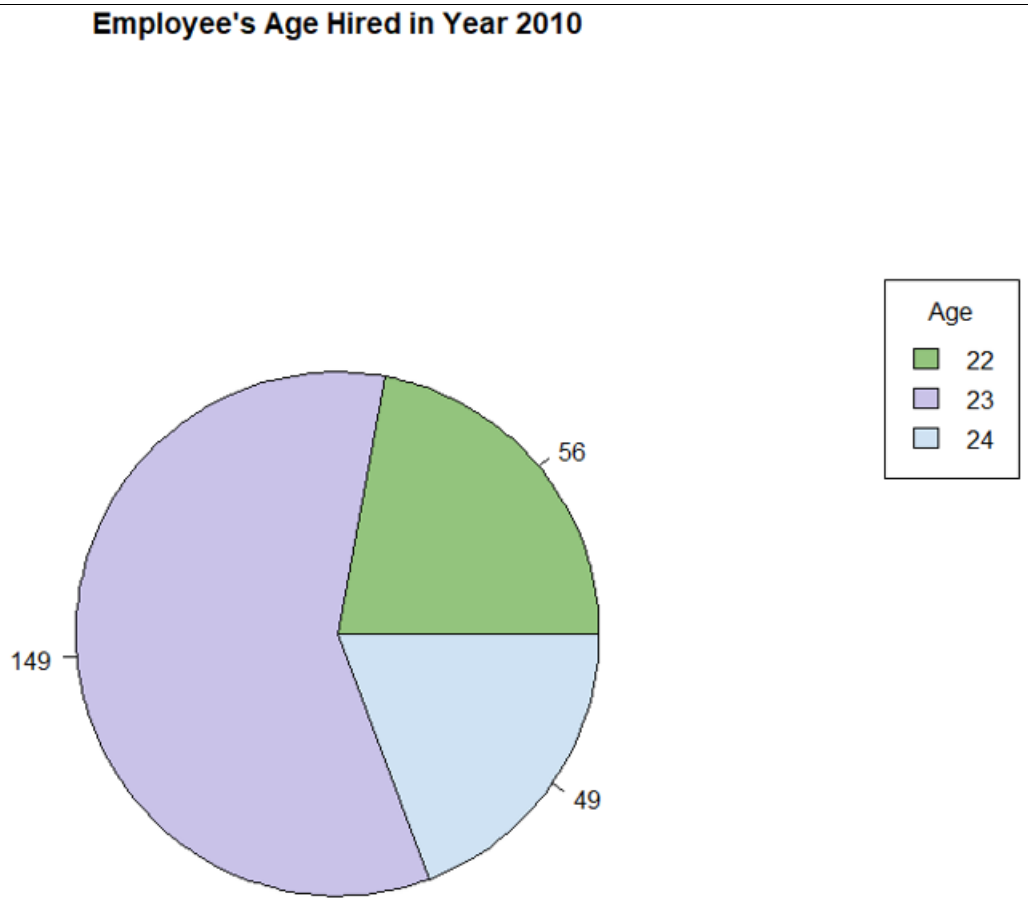
To plot facet in R programming language, we use the facet_warp () function from the ggplot2 library. The facet_wrap () is used to form a matrix of panels defined by row and column faceting variables, to display all combinations of the variables that exist in the data. By using the facet_wrap() functions, it helps display the information on specific requirement to help audience to see a clear data.

Colour Palette

Source code

```
#in 2010
hire_2010 = table(data2 %>% filter(Year_of_Status=="2010",Hire_Year=="2010") %>% select(Age))
pie(hire_2010,label=hire_2010,main="Employee's Age Hired in Year 2010",col=c("#93c47d","#c9c2e8","#cfe2f3"),border="black")
legend("topright",c("22","23","24"), cex =1, fill=c("#93c47d","#c9c2e8","#cfe2f3"),title="Age")
```

Output



Explanation

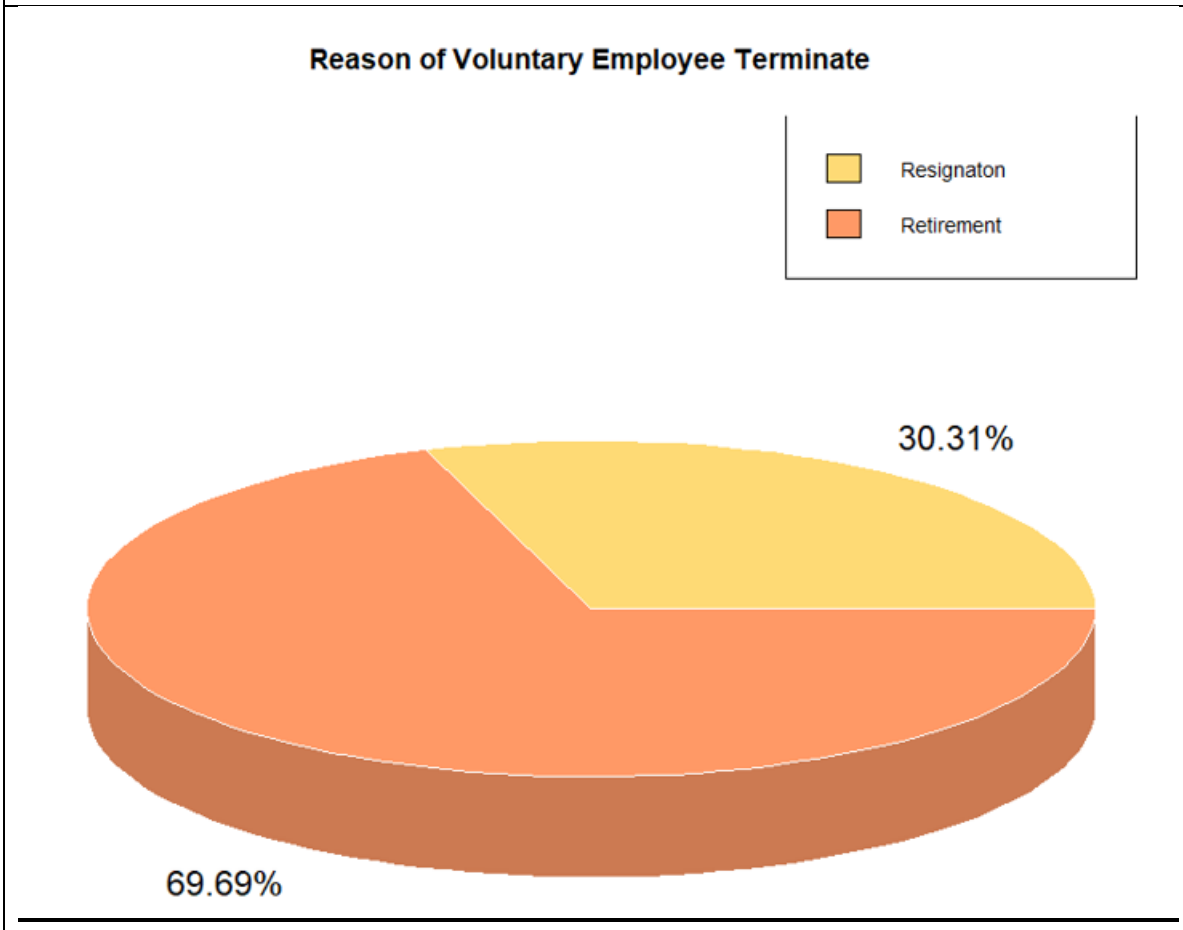
R colour palettes for changing the default colour of a graph generated using either the ggplot2 package or the R base plot functions. By using the col-c() functions, it helps display the information on specific requirement to fill in certain colours to help audience to differentiate data.

Legend

Source code

```
#reason
r_volun= table(data2 %>% filter(Status =="TERMINATED" & Type_of_Termination=="voluntary") %>% select(Reason_of_Termination))
rvPercent = round(100*r_volun/sum(r_volun),2)
pie3D(r_volun,labels=paste0(rvPercent,"%"),cex=0.7,radius=1,main="Reason of Voluntary Employee Terminate",border="white", col=c("#feda75","#ff9966"))
legend("topright",c("Resignaton","Retirement"),cex=0.9,fill=c("#feda75","#ff9966"))
```

Output



Explanation

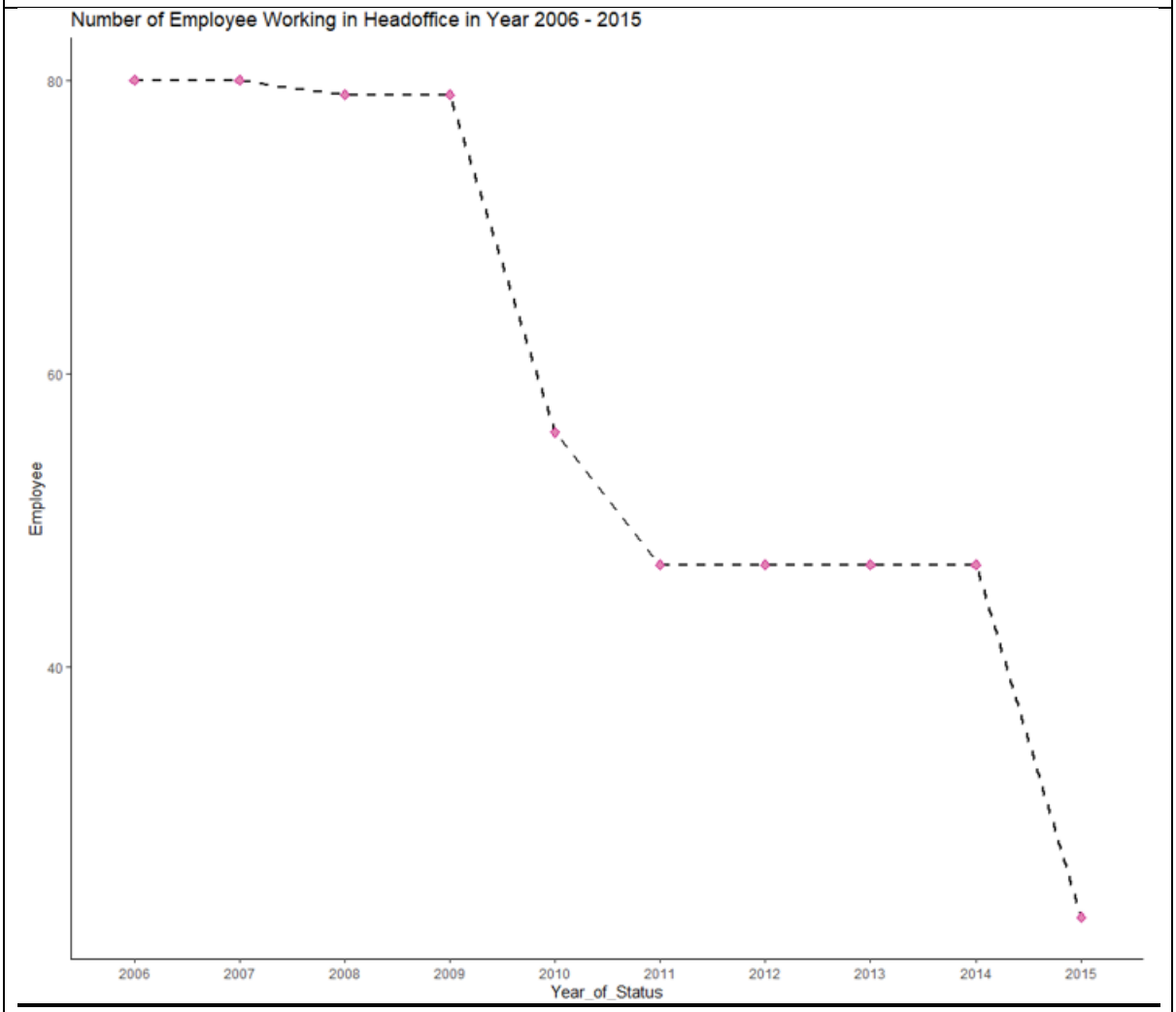
Legend function in R adds legend box to the plot. legend () function in R makes graph easier to read and interpret in better way. By using the legend() functions, it helps display the information on a legend box to help audience to see a clear data by labelling it.

Line Type

Source code

```
#Headoffice
data2 %>% filter(Business_Unit=="HEADOFFICE")%>%group_by(Year_of_Status,Business_Unit) %>%summarise(Employee=n()) %>% ggplot(aes(x=Year_of_Status,y=Employee,group=Business_Unit))+
  geom_line(color="black", size=1, alpha=0.9, linetype=2)+geom_point(shape=23, color="#c71585", fill="#e485b4", size=2)+
  ggtitle("Number of Employee working in Headoffice in Year 2006 - 2015")+theme_classic()
```

Output



Explanation

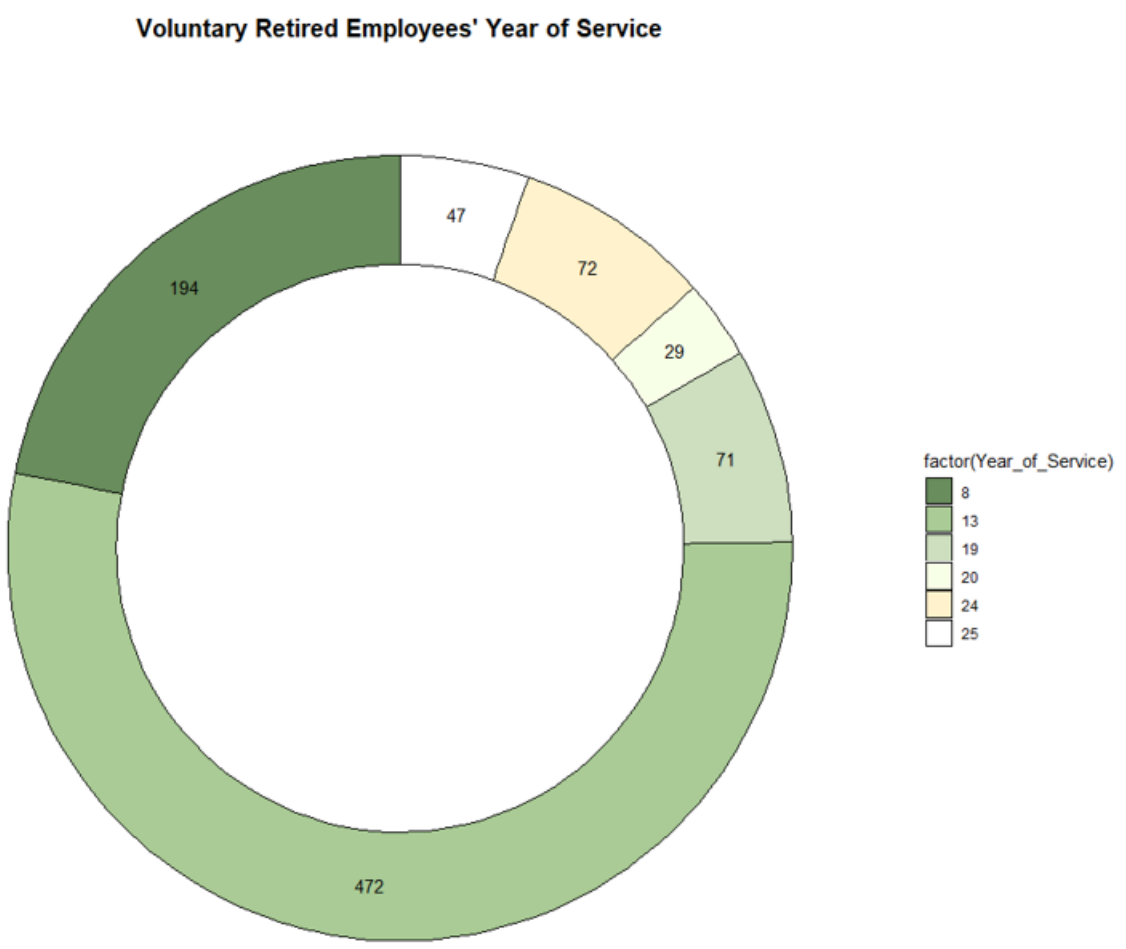
The lines () function is part of the R graphics package, and it's used to add lines to the plot. By using the geon_line() functions, it helps display the information on specific requirement to help audience to see the data visualization.

Doughnut Chart

Source code

```
#length
data2 %>% filter(Status =="TERMINATED" & Type_of_Termination=="Voluntary" & Reason_of_Termination=="Retirement") %>%
  group_by(Year_of_Service) %>% summarise(Employee=n()) %>% ggplot(aes(x=3,y=Employee,fill=factor(Year_of_Service)))+
  geom_col(col="black")+coord_polar(theta="y")+ xlim(c(0.2,3.5))+theme(panel.background = element_rect(fill = "white"),
                        axis.title = element_blank(),axis.ticks = element_blank(),axis.text = element_blank())+
  geom_text(aes(label=Employee),position = position_stack(vjust =0.5))+
  ggtitle("Voluntary Retired Employees' Year of Service")+
  scale_fill_manual(values=c("#698d5d","#aacb96","#cddfbf","#f8ffe7","#fff2cc","white"))+
  theme(plot.title=element_text(size=15,face="bold",hjust = 0.5))
```

Output



Explanation

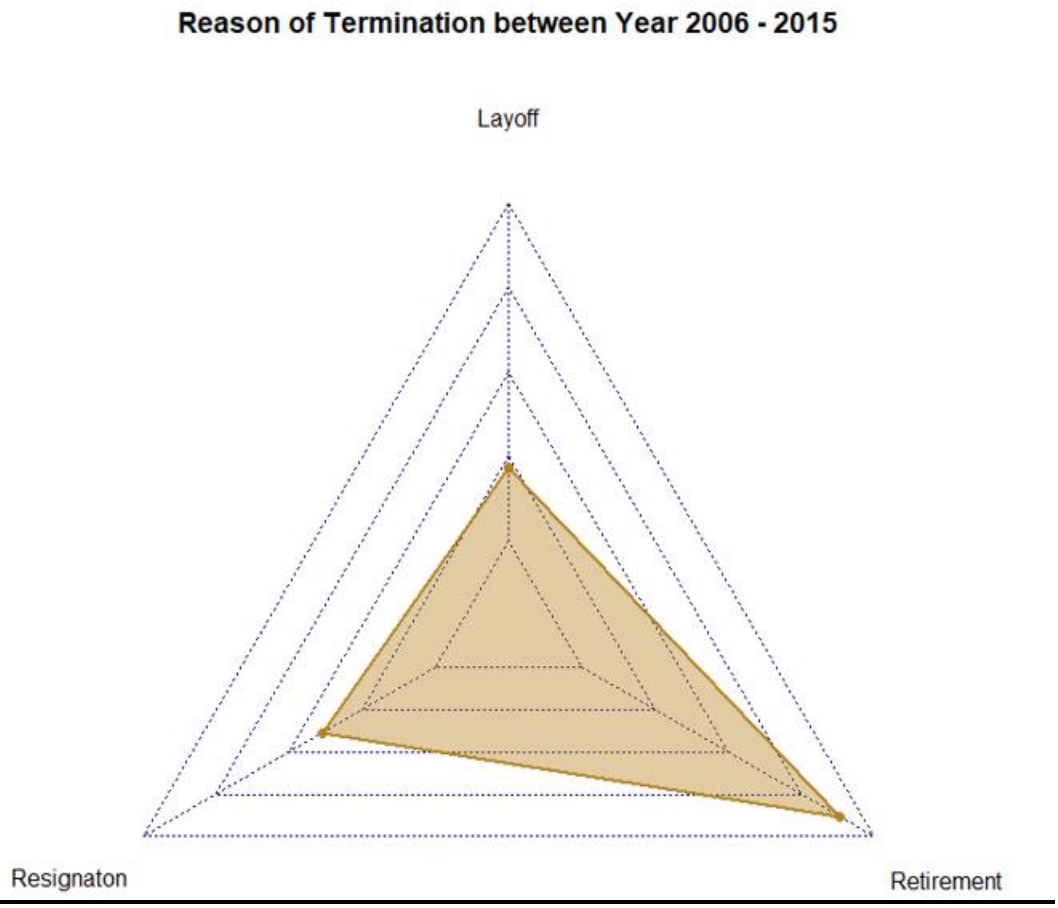
Donut or doughnut charts are an alternative chart for pie charts, which have a hole in the middle, making them cleaner to read than pie charts. By using the donut chart, it helps display the information on specific requirement to help audience to visualize the data.

Radar Chart

Source code

```
#reason
layoff_t = matrix(rbind(max=1000,min=0,values=as.integer(data2 %>% filter( Reason_of_Termination=="Layoff") %>%
                                                         summarise(Layoff=n())))
resign_t = matrix(rbind(max=1000,min=0,values=as.integer(data2 %>% filter( Reason_of_Termination=="Resignaton") %>%
                                                         summarise(Layoff=n()))))
retire_t = matrix(rbind(max=1000,min=0,values=as.integer(data2 %>% filter( Reason_of_Termination=="Retirement") %>%
                                                         summarise(Layoff=n()))))

terminate_emp = data.frame(cbind(layoff_t,resign_t,retire_t))
radarchart(terminate_emp,title = "Reason of Termination between Year 2006 - 2015",
           vlabels = c("Layoff","Resignaton","Retirement"),
           pcol=rgb(0.7,0.5,0.1,0.9) , pfcol= rgb(0.7,0.5,0.1,0.4) ,plwd=2)
```

Output



Explanation

A radar chart, also known as a spider plot, is used to visualize the values or scores assigned to an individual over multiple quantitative variables, where each variable corresponds to a specific axis**.** By using the radar graph, it helps display the information on specific requirement to help audience to visualize the data.

Conclusion

Through doing the analysis above, the company could act in maintaining their employee retention while focusing on certain factors during their hiring. The company human resource could also have proper preparations to counter such problems. They may know which age group, cities and gender will stay longer in a company whilst avoiding employee who might be terminated at a young age.

Based on the analysis above, we could also see the hiring and firing trend of the company. We could see that the company have a specific date of both hiring and firing thus employee could focus more on specific months of the year to avoid termination.

Reference

Letter, F. and Low?, W., 2022. *San Francisco Fed*. [online] San Francisco Fed. Available at: <https://www.frbsf.org/economic-research/publications/economic-letter/2020/march/why-is-unemployment-currently-so-low/> [Accessed 21 August 2022].

BrainStation® Blog. 2022. *Why Is Gender Balance so Important for Companies (and How Can They Improve It)?*. [online] Available at: <https://brainstation.io/blog/why-is-gender-balance-so-important-for-companies-and-how-can-they-improve-it> [Accessed 21 August 2022].

Chan, &. and Chan, L., 2022. *Why do people retire?*. [online] Nestegg.com.au. Available at: <https://www.nestegg.com.au/retirement/planning/why-do-people-retire#:~:text=The%20second%20most%20common%20reason%20for%20retirement%20is,other%20circumstances%2C%20regardless%20of%20their%20intention%20to%20stay.> [Accessed 21 August 2022].

Indeed Career Guide. 2022. *Employee Performance Evaluation: Definition, Benefits and How To Prepare | Indeed.com*. [online] Available at: <https://www.indeed.com/career-advice/career-development/employee-performance-evaluation> [Accessed 21 August 2022].

Wealth Awesome. 2022. *Best Age to Retire in Canada: 55, 65, or Never?*. [online] Available at: <https://wealthawesome.com/best-age-to-retire-in-canada/#:~:text=%20Best%20Age%20to%20Retire%20in%20Canada%3A%2055%2C,Many%20Canadians%20dream%20about%20retiring%20early%2C...%20More%20> [Accessed 21 August 2022].

Bc.edu. 2022. *Why do women retire earlier than men?*. [online] Available at: <https://www.bc.edu/bc-web/schools/ssw/sites/center-on-aging-and-work/news-and-notes/blog/why-do-women-retire-earlier-than-men-.html> [Accessed 21 August 2022].

Stlawyers.ca. 2022. [online] Available at: <https://stlawyers.ca/blog-news/increasing-costs-of-terminating-older-employees/> [Accessed 21 August 2022].

Olaru, M., 2022. *11 Top Reasons Why Talented Millennial Employees Get Fired*. [online] AlignThoughts. Available at: <https://alignthoughts.com/10-reasons-why-talented-millennial-employees-get-fired/> [Accessed 21 August 2022].

Winters, E. and Winters, E., 2022. *Termination Due to Lack of Desired Skills Not Age Bias*. [online] SHRM. Available at: <https://www.shrm.org/resourcesandtools/legal-and-compliance/employment-law/pages/termination-desired-skills.aspx> [Accessed 21 August 2022].

Time. 2022. *Young People Are Leaving Their Jobs in Record Numbers—And Not Going Back*. [online] Available at: <https://time.com/6111245/young-workers-quitting/> [Accessed 21 August 2022].

2022. [online] Available at: <https://www.statcan.gc.ca/en/dai/smr08/2017/smr08_216_2017> [Accessed 21 August 2022].

The Balance Careers. 2022. *What Are the Key Causes of Employment Termination?*. [online] Available at: <https://www.thebalancecareers.com/what-causes-employment-termination-1918275> [Accessed 21 August 2022].