

Consultancy Project

Tara Dehdari

2023-05-16

Pre-processing Data

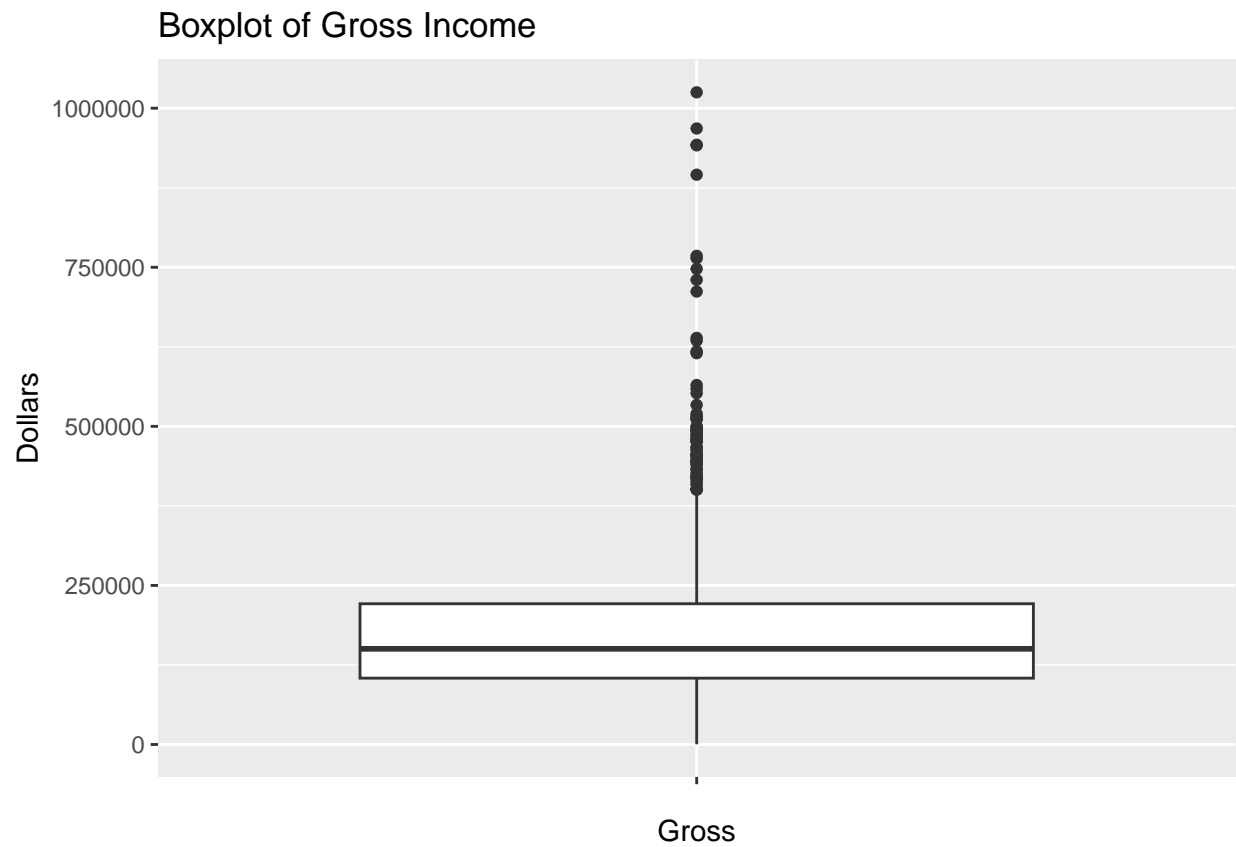
Missing values

```
missing_counts <- colSums(is.na(Employee_monthly_salary))
print(missing_counts)
```

```
##           EmpID           Name           Gender
##           0           0           0
##   Date_of_Birth           Age           Join_Date
##           0           0           0
## Tenure_in_org_in_months           GROSS           Net_Pay
##           0           0           0
##           Deduction Deduction_percentage           Designation
##           0           0           0
##           Department
##           0
```

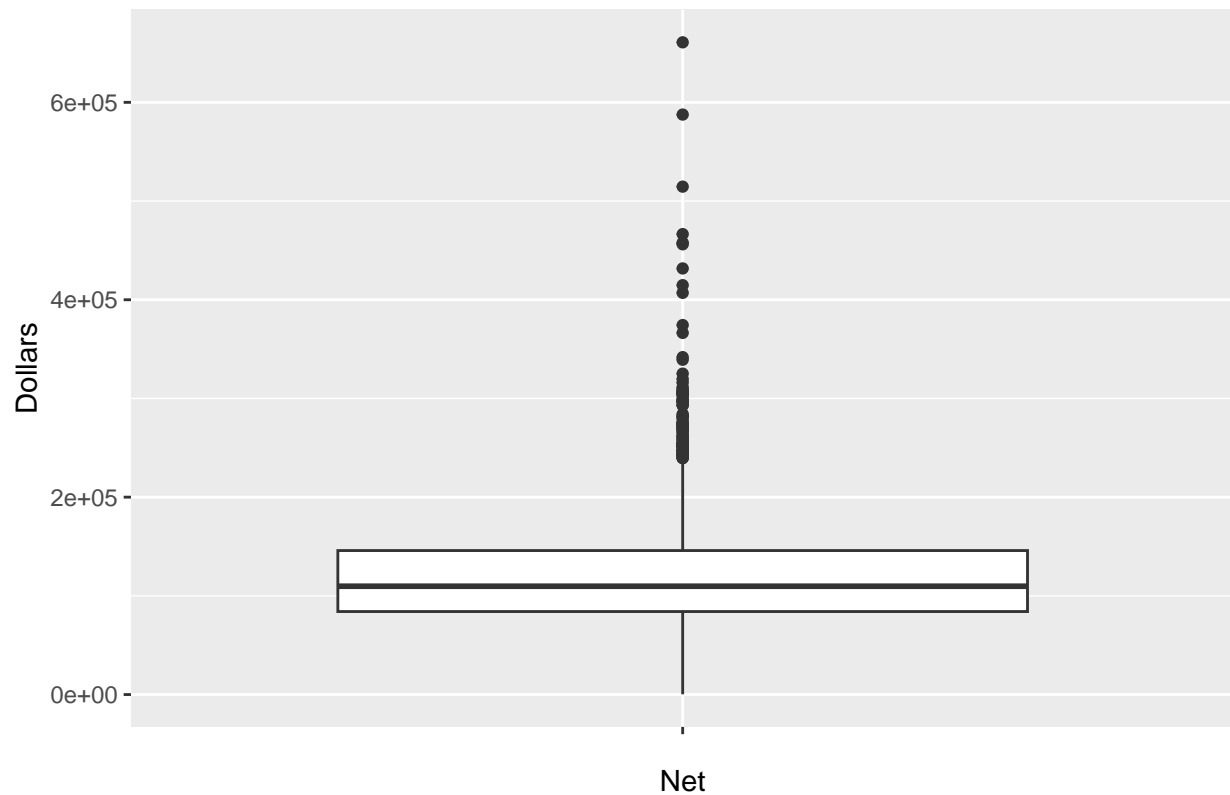
Looking for Outliers

```
ggplot(Employee_monthly_salary, aes(x = "", y = GROSS)) +
  geom_boxplot() +
  labs(title = "Boxplot of Gross Income",
       x = "Gross",
       y = "Dollars")
```



```
ggplot(Employee_monthly_salary, aes(x = "", y = Net_Pay)) +  
  geom_boxplot() +  
  labs(title = "Boxplot of Net Income",  
        x = "Net",  
        y = "Dollars")
```

Boxplot of Net Income



Data Description

```
str(Employee_monthly_salary)
```

```
## spc_tbl_ [1,802 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ EmpID           : num [1:1802] 19575 19944 20055 20058 20332 ...
##  $ Name            : chr [1:1802] "Keven Norman" "Kristin Werner" "Avery Barber" "Boris Gibson" ...
##  $ Gender          : chr [1:1802] "M" "F" "M" "M" ...
##  $ Date_of_Birth   : chr [1:1802] "3/9/94" "23/06/1994" "27/02/1996" "29/09/1993" ...
##  $ Age             : num [1:1802] 25 26 24 26 29 32 27 24 24 24 ...
##  $ Join_Date       : chr [1:1802] "2/12/19" "13/01/2020" "14/11/2019" "13/01/2020" ...
##  $ Tenure_in_org_in_months: num [1:1802] 7 6 8 6 25 8 6 6 8 6 ...
##  $ GROSS           : num [1:1802] 74922 44375 82263 44375 235405 ...
##  $ Net_Pay         : num [1:1802] 71494 39971 77705 40164 143963 ...
##  $ Deduction       : num [1:1802] 3428 4404 4558 4211 91442 ...
##  $ Deduction_percentage : num [1:1802] 4.58 9.92 5.54 9.49 38.84 ...
##  $ Designation     : chr [1:1802] "Product Operations Analyst.Associate." "Platform Operations" ...
##  $ Department      : chr [1:1802] "IT Product Management & Ops" "Platform Operations" "Platform Operations" ...
##  - attr(*, "spec")=
##    .. cols(
##    ..   EmpID = col_double(),
##    ..   Name = col_character(),
##    ..   Gender = col_character(),
##    ..   Date_of_Birth = col_character(),
```

```
## .. Age = col_double(),
## .. Join_Date = col_character(),
## .. Tenure_in_org_in_months = col_double(),
## .. GROSS = col_double(),
## .. Net_Pay = col_double(),
## .. Deduction = col_double(),
## .. Deduction_percentage = col_double(),
## .. Designation = col_character(),
## .. Department = col_character()
## .. )
## - attr(*, "problems")=<externalptr>
```

```
num_records <- nrow(Employee_monthly_salary)
num_fields <- ncol(Employee_monthly_salary)

cat("Number of records:", num_records, "\n")
```

```
## Number of records: 1802
```

```
cat("Number of fields:", num_fields, "\n")
```

```
## Number of fields: 13
```

```
#number of females
num_females <- sum(Employee_monthly_salary$Gender == "F")
#number of males
num_males <- sum(Employee_monthly_salary$Gender == "M")
#average salary of females
avg_F_salary <- Employee_monthly_salary %>%
  filter(Gender == "F") %>%
  summarize(avg_F_salary = mean(GROSS))
#average salary of males
avg_M_salary <- Employee_monthly_salary %>%
  filter(Gender == "M") %>%
  summarize(avg_M_salary = mean(GROSS))
```

```
cat("The number of Females are: ", num_females, " and the average gross pay for females is:" , avg_F_salary)
```

```
## The number of Females are: 499 and the average gross pay for females is: 140560.3
```

```
cat("The number of Males are: ", num_males, " and the average gross pay for males is:" , avg_M_salary)
```

```
## The number of Males are: 1303 and the average gross pay for males is: 189332.3
```

```
#age range
age_range <- range(Employee_monthly_salary$Age)

cat("Age range:", age_range[1], "-", age_range[2], "\n")
```

```
## Age range: 21 - 56
```

```
#number of each person in each age
age_counts <- table(Employee_monthly_salary$Age)

print(age_counts)
```

```
##
##  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35  36  37  38  39  40
##    2  23  39  48  80 109 112 126 170 155 133 136 105  89  87  81  70  53  43  35
##  41  42  43  44  45  46  47  48  49  50  51  53  55  56
##  29  28  11  11   3   7   6   1   2   4   1   1   1   1
```

Data Exploration

```
#how many jobs
unique_jobs <- unique(Employee_monthly_salary$Designation)
print(unique_jobs)
```

```
## [1] "Product Operations Analyst.Associate."
## [2] "Platform Operations Engineer.Associate."
## [3] "Software Engineer.Senior."
## [4] "LMS Administrator II.."
## [5] "Technical Solutions Engineer.Associate."
## [6] "Technical Solutions Engineer.."
## [7] "Account Service Manager.."
## [8] "Network Administrator.."
## [9] "Business Systems Analyst.Senior."
## [10] "Client Service Manager.."
## [11] "Business Operations Analyst.Senior."
## [12] "Systems Support Engineer II.."
## [13] "Server Tech Services Engineer.."
## [14] "Data Scientist.Senior."
## [15] "Software Engineer.Senior II."
## [16] "Software Development Engineer in Test.Senior."
## [17] "Software Engineer II.."
## [18] "Support Delivery Manager.."
## [19] "Global Corporate Card Specialist.Senior."
## [20] "Manager.Senior.Financial Operations"
## [21] "Platform Operations Engineer.."
## [22] "Manager.Senior.Engineering"
## [23] "Security Operations Center Specialist II.."
## [24] "Software Engineer.."
## [25] "Release Engineer II.."
## [26] "System Operations Engineer.Senior."
## [27] "Business Systems Analyst.Senior II."
## [28] "Q/A Engineer.Senior."
## [29] "Program Manager.."
## [30] "Solutions Architect II.."
## [31] "Site Reliability Engineer.Senior."
## [32] "Infrastructure Delivery Manager.."
## [33] "Business Operations Analyst II.."
## [34] "Systems Support Engineer.."
```

```

## [35] "Program Manager.Senior.Technical"
## [36] "Software Development Engineer in Test II.."
## [37] "Applications Developer II.."
## [38] "Technical Project Manager II.."
## [39] "Systems Administrator..Windows"
## [40] "Systems Engineer.Senior."
## [41] "Business Operations Analyst.."
## [42] "Contract/Billing Analyst.."
## [43] "Solutions Architect.Senior."
## [44] "Systems Engineer II.."
## [45] "Solutions Architect.."
## [46] "Media Operations Engineer.Associate."
## [47] "Product Manager.Senior."
## [48] "Network Administrator II.."
## [49] "Finance Systems Analyst.Senior Lead."
## [50] "Order Analyst.Senior."
## [51] "Security Operations Center Specialist.Senior II."
## [52] "Technical Account Manager.Senior."
## [53] "Relationship Manager.Senior."
## [54] "Instructional Designer.Senior."
## [55] "Contract/Billing Analyst II.."
## [56] "User Interface Engineer II.."
## [57] "Client Service Manager.Senior II."
## [58] "Software Development Engineer in Test.Senior II."
## [59] "Collections Specialist.Senior."
## [60] "Manager..Engineering"
## [61] "Software Engineer.Senior II Lead."
## [62] "Business Systems Analyst.Senior II Lead."
## [63] "Program Manager.Senior."
## [64] "Q/A Engineer II.."
## [65] "Technical Account Manager.."
## [66] "Marketing Program Specialist.."
## [67] "Software Development Engineer in Test.Senior II Lead."
## [68] "Channel Support Manager.."
## [69] "Performance Engineer.Senior II."
## [70] "Program Manager..Technical"
## [71] "Security Analyst.Senior."
## [72] "Platform Operations Engineer.Senior."
## [73] "Q/A Engineer.Senior II."
## [74] "Systems Support Engineer.Senior."
## [75] "Client Service Manager.Senior."
## [76] "Manager..Q/A Engineering"
## [77] "Network Infrastructure Engineer II.."
## [78] "Manager..Program Management Technical"
## [79] "Supervisor..Help Desk"
## [80] "Corporate Services Specialist.Senior Lead."
## [81] "Network Engagement Consultant.."
## [82] "Release Engineer.Senior."
## [83] "Commissions Analyst II.."
## [84] "Lead..Account Management"
## [85] "Network Implementation Manager II.Lead."
## [86] "Partner Services Manager.."
## [87] "LMS Administrator.Senior Lead."
## [88] "Systems Administrator.Senior.UNIX"

```

```

## [89] "Platform Operations Engineer II.."
## [90] "Graphic Designer.Senior Lead."
## [91] "Manager.Senior.Software Development Engineer in Test"
## [92] "Systems Administrator II.."
## [93] "Partner Services Manager.Senior."
## [94] "Site Reliability Engineer II.."
## [95] "Financial Analyst.Senior."
## [96] "Contract/Billing Analyst.Senior."
## [97] "Solutions Engineer II.."
## [98] "Booking Operations Analyst II.."
## [99] "Practice Manager.Senior."
## [100] "Collections Specialist.."
## [101] "Accountant II.."
## [102] "Program Manager II..Technical"
## [103] "Program Manager II.."
## [104] "Server Tech Services Engineer II.Lead."
## [105] "Architect.Senior."
## [106] "Solutions Architect.Senior Lead."
## [107] "Performance Engineer II.."
## [108] "Manager..Software Development Engineer in Test"
## [109] "Billing Specialist.Senior."
## [110] "Architect.."
## [111] "Recruiter II.."
## [112] "Systems Support Engineer.Senior Lead."
## [113] "Q/A Engineer.Principal."
## [114] "Technical Solutions Engineer II.Lead."
## [115] "Marketing Manager.Senior."
## [116] "Applications Developer.Senior."
## [117] "Technical Solutions Engineer II.."
## [118] "Systems Administrator II..UNIX"
## [119] "Director..Service Line"
## [120] "Enterprise Security Architect.Senior II."
## [121] "Database Administrator.Senior Lead."
## [122] "Manager..Network Design"
## [123] "Software Development Engineer in Test.Senior Lead."
## [124] "Order Analyst II.."
## [125] "Documentation and Training Specialist.Senior."
## [126] "Program Manager.Associate."
## [127] "User Interface Engineer.Senior II."
## [128] "Manager..Business Operations"
## [129] "Manager.Senior.Solutions Engineering"
## [130] "Manager..Sales Operations"
## [131] "Order Analyst.Senior Lead."
## [132] "Manager..Network Infrastructure"
## [133] "Network Engineer II.."
## [134] "Practice Manager.."
## [135] "Manager..Reliability Engineering"
## [136] "Support Delivery Manager II.."
## [137] "Site Reliability Engineer.Senior II."
## [138] "Systems Administrator.Senior.Windows"
## [139] "Manager..Account Management"
## [140] "Manager..Technical Support"
## [141] "Manager..Travel and Vendor Management"
## [142] "Director..Engineering"

```

[143] "Program Manager.Senior Lead."
 ## [144] "Escalation Manager.."
 ## [145] "Manager..Platform Operations"
 ## [146] "Technical Solutions Engineer.Senior."
 ## [147] "Director..Accounting"
 ## [148] "Technical Solutions Engineer.Senior Lead."
 ## [149] "Systems Support Engineer.Senior II Lead."
 ## [150] "Deal Desk Analyst.Senior."
 ## [151] "Platform Operations Engineer.Senior Lead."
 ## [152] "ECG Sales Manager.."
 ## [153] "Billing Specialist.Senior Lead."
 ## [154] "Manager.Senior.Service Line"
 ## [155] "Director.Senior.Technical Support"
 ## [156] "Services Analyst.Principal."
 ## [157] "System Operations Engineer.Senior Lead."
 ## [158] "Network Engagement Consultant II.."
 ## [159] "Manager..ECG"
 ## [160] "Manager.Senior.Network Delivery Architect"
 ## [161] "Strategic Sales Director.."
 ## [162] "Director..Platform Operations"
 ## [163] "Manager.Senior.Compensation and Benefits"
 ## [164] "Billing Specialist.."
 ## [165] "Finance Systems Analyst.."
 ## [166] "Manager.Senior.Learning & Development"
 ## [167] "User Interface Engineer.Senior."
 ## [168] "Q/A Engineer.Senior II Lead."
 ## [169] "Software Engineer.Principal."
 ## [170] "Revenue Accountant.Senior Lead."
 ## [171] "Security Operations Center Specialist.."
 ## [172] "Compensation and Benefits Specialist.Senior."
 ## [173] "Network Implementation Manager.."
 ## [174] "Network Design Engineer II.."
 ## [175] "Relationship Manager.."
 ## [176] "Manager.Senior.Talent Sourcing and Research"
 ## [177] "Software Development Engineer in Test.Principal Lead."
 ## [178] "Network Support Consultant.."
 ## [179] "Media Operations Engineer.Senior."
 ## [180] "Manager.Senior.Technical Support"
 ## [181] "Contract/Billing Analyst.Senior Lead."
 ## [182] "System Operations Engineer.Senior II Lead."
 ## [183] "Software Development Engineer in Test.Principal."
 ## [184] "Commissions Analyst.Senior."
 ## [185] "Order Management Supervisor.."
 ## [186] "Manager..Product Operations"
 ## [187] "Solutions Architect.Senior II."
 ## [188] "Vice President..Platform Delivery"
 ## [189] "Graphic Designer II.."
 ## [190] "Manager.Senior.Q/A Engineering"
 ## [191] "Manager.Senior.Performance Engineering"
 ## [192] "Product Operations Analyst.Senior."
 ## [193] "Office Manager.."
 ## [194] "Marketing Automation Specialist.Senior."
 ## [195] "Financial Analyst II.."
 ## [196] "Manager..Finance"

[197] "Relationship Manager II.."
 ## [198] "Shift Manager..Platform Operations"
 ## [199] "Director..Technical Support"
 ## [200] "Manager..Contract Management"
 ## [201] "Manager..Booking Operations"
 ## [202] "Security Operations Center Specialist.Senior."
 ## [203] "Marketing Program Specialist.Senior."
 ## [204] "Manager.Senior.IT Help Desk Services"
 ## [205] "Manager..System Operations"
 ## [206] "Logistics Specialist.Senior."
 ## [207] "Collections Supervisor.."
 ## [208] "Product Support Engineer.Senior."
 ## [209] "Technical Account Manager II.."
 ## [210] "Order Analyst.."
 ## [211] "Partner Enablement Manager.Senior."
 ## [212] "Deal Desk Analyst.Senior Lead."
 ## [213] "Logistics Analyst.Lead."
 ## [214] "Software Engineer.Principal Lead."
 ## [215] "Director..Account Management"
 ## [216] "Manager..Training"
 ## [217] "Manager.Senior.Learning Solutions"
 ## [218] "Human Resources Business Partner II.."
 ## [219] "Manager.Senior.Digital Properties"
 ## [220] "Revenue Accountant.."
 ## [221] "Director..Customer Financial Operations"
 ## [222] "Technical Project Manager.Senior."
 ## [223] "Systems Support Engineer.Senior II."
 ## [224] "Recruiter.Senior Lead."
 ## [225] "Network Infrastructure Engineer.Senior Lead."
 ## [226] "Accounts Receivable Specialist.Senior."
 ## [227] "Hardware Performance Engineer II.."
 ## [228] "Business Operations Analyst.Senior Lead."
 ## [229] "Director..Program Management Technical"
 ## [230] "Product Operations Analyst II.."
 ## [231] "Help Desk Specialist.."
 ## [232] "Payroll Analyst.Senior."
 ## [233] "Talent Sourcing and Research Specialist II.."
 ## [234] "Services Analyst II.."
 ## [235] "Web Developer.Senior."
 ## [236] "Marketing Program Manager.."
 ## [237] "Performance Engineer.Senior II Lead."
 ## [238] "Network Delivery Architect II.."
 ## [239] "Manager..Online Properties"
 ## [240] "Network Engineer.Senior Lead."
 ## [241] "Customer Incident Manager II.."
 ## [242] "Manager.Senior.Network Service Management"
 ## [243] "Incident Manager.."
 ## [244] "Help Desk Administrator.."
 ## [245] "Collections Specialist II.."
 ## [246] "Product Support Engineer.."
 ## [247] "Network Delivery Architect.Senior."
 ## [248] "Systems Administrator.Senior Lead.UNIX"
 ## [249] "Director..Finance"
 ## [250] "Accountant.Senior."

```

## [251] "Release Engineer.Senior Lead."
## [252] "Director.Senior.Engineering"
## [253] "Database Administrator II.."
## [254] "Director..Program Management"
## [255] "Customer Incident Manager.."
## [256] "Booking Operations Analyst.Senior Lead."
## [257] "Manager.Senior.Recruiting"
## [258] "Network Implementation Manager Ii.."
## [259] "Manager..Service Line"
## [260] "Recruiter.Senior."
## [261] "Collections Specialist.Senior Lead."
## [262] "Counsel.."
## [263] "Data Engineer.Senior."
## [264] "Vice President & Managing Director India.."
## [265] "Manager.Senior.ECG"
## [266] "Manager.Senior.Account Management"
## [267] "Services Analyst.Senior."
## [268] "Manager.Senior.Product Operations"
## [269] "Security Operations Center Network Engineer.."
## [270] "Manager..Network Administration"
## [271] "Systems Engineer.Principal."
## [272] "Revenue Accountant II.."
## [273] "Executive Assistant.."
## [274] "Technical Enablement Strategist.."
## [275] "Applications Developer.Senior Lead."
## [276] "Platform Operations Engineer II.Lead."
## [277] "Technical Project Manager.Senior Lead."
## [278] "Manager..Enterprise Security Operations"
## [279] "Manager.Senior.Marketing"
## [280] "System Operations Engineer.Senior II."
## [281] "Accountant.Senior Lead."
## [282] "Manager.Senior.Platform Operations"
## [283] "Reliability Engineer.Senior."
## [284] "Manager.Senior.Systems Support"
## [285] "Manager..Order Management"
## [286] "Human Resources Business Partner.Senior."
## [287] "Solutions Engineer.Senior."
## [288] "Support Delivery Manager.Senior."
## [289] "Manager..Systems Engineering"
## [290] "Manager..Collections"
## [291] "Performance Engineer.Senior."
## [292] "Network Engineer.."
## [293] "Product Architect.."
## [294] "Web Content Publisher.Senior."
## [295] "Manager.Senior.Technical Enablement"
## [296] "Director..Product Operations"
## [297] "Infrastructure Delivery Manager II.."
## [298] "Systems Administrator.Senior Lead.Windows"
## [299] "Manager.Senior.Business Applications"
## [300] "Network Infrastructure Engineer.."
## [301] "Deal Desk Analyst II.."
## [302] "Product Marketing Manager.."
## [303] "Business Systems Analyst.."
## [304] "Program Manager.Principal.Technical"

```

```

## [305] "Web Content Publisher.."
## [306] "Director..Human Resources"
## [307] "Manager.Senior.Sales Operations"
## [308] "Process & Quality Engineer.Senior Lead."
## [309] "Web Developer.Senior Lead."
## [310] "Technical Enablement Strategist.Senior."
## [311] "Manager..HR Operations"
## [312] "Booking Operations Analyst.."
## [313] "Manager.Senior.Business Operations"
## [314] "Manager.Senior.Systems Administration"
## [315] "Payroll Analyst.."
## [316] "Systems Administrator II..Windows"
## [317] "Documentation and Training Specialist.."
## [318] "Technical Writer.Senior II."
## [319] "Solutions Architect II.Lead."
## [320] "Web Developer II.."
## [321] "Director..Product Line"
## [322] "Financial Analyst.Senior Lead."
## [323] "Architect.Lead."
## [324] "Technical Consultant.."
## [325] "Capacity Planning Manager.."
## [326] "Human Resources Specialist II.."
## [327] "Engagement Manager II.."
## [328] "Product Support Engineer.Senior Lead."
## [329] "Manager..Payroll"
## [330] "Network Account Manager.."
## [331] "Systems Administrator.."
## [332] "Security Operations Technician.Senior."
## [333] "Security Operations Technician II.."
## [334] "Developer Evangelist.."
## [335] "Media Operations Engineer.."
## [336] "System Operations Engineer.."
## [337] "Client Service Manager.Associate."
## [338] "Software Development Engineer in Test.."
## [339] "Data Engineer II.."
## [340] "Media Operations Engineer II.."
## [341] "Support Delivery Specialist.."
## [342] "Data Engineer.Senior II Lead."
## [343] "Product Operations Analyst.."
## [344] "Software Engineer.Senior Lead."
## [345] "Systems Engineer.."
## [346] "Learning & Development Specialist II.Lead."
## [347] "Relationship Manager.Senior Lead."
## [348] "Network Engineer.Senior."
## [349] "Network Design Engineer.."
## [350] "Architect.Principal Lead."
## [351] "Logistics Specialist.."
## [352] "Systems Engineer.Associate."
## [353] "Travel and Vendor Management Specialist.."
## [354] "Product Marketing Specialist.Associate."
## [355] "Channel Marketing Specialist.Senior."
## [356] "Service Architect II.."
## [357] "Data Engineer.."
## [358] "Server Tech Services Engineer II.."

```

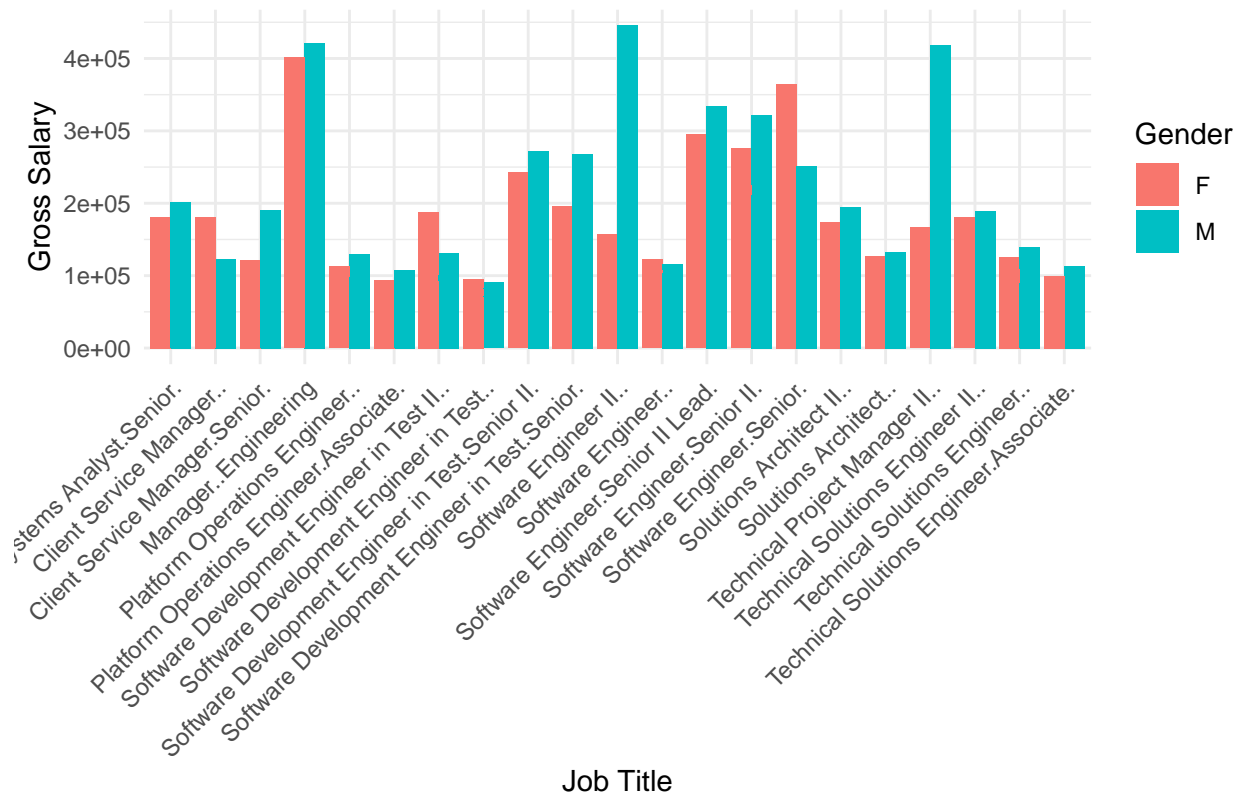
```
## [359] "Billing Specialist II.."
## [360] "Help Desk Specialist.Associate."
## [361] "Manager..Solutions Engineering"
## [362] "Booking Operations Analyst.Senior."
## [363] "Security Operations Center Specialist.Senior II Lead."
## [364] "Channel Support Manager.Senior."
## [365] "Enterprise Sales Executive.Principal.New and Existing Accts"
## [366] "Lab Manager II.."
## [367] "Telecom Engineer II.."
## [368] "Technical Consultant II.."
## [369] "Enterprise Security Architect.Senior."
## [370] "Site Reliability Engineer.."
## [371] "Architect.Senior Lead."
## [372] "Help Desk Administrator.Senior."
## [373] "Finance Systems Analyst.Associate."
## [374] "Account Service Manager.Associate."
## [375] "Q/A Engineer.."
## [376] "Sales Development Representative.."
## [377] "System Operations Engineer II.."
## [378] "Security Operations Technician.."
## [379] "Process & Quality Engineer.."
## [380] "Learning & Development Specialist II.."
## [381] "Financial Analyst.."
## [382] "Solutions Architect.Associate."
## [383] "Technical Support Engineer.."
## [384] "Manager.Senior.Facilities Operations"
## [385] "Engagement Manager.Senior."
## [386] "User Interface Engineer.."
## [387] "Counsel.Senior."
## [388] "Performance Engineer.."
## [389] "Contract/Billing Analyst.Associate."
## [390] "Data Scientist.."
```

```
#filter by top job titles
top_job_titles <- Employee_monthly_salary %>%
  count(Designation) %>%
  top_n(20)
```

```
## Selecting by n
```

```
#barchart with filtered job titles
ggplot(filter(Employee_monthly_salary, Designation %in% top_job_titles$Designation),
  aes(x = Designation, y = GROSS, fill = Gender)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(x = "Job Title", y = "Gross Salary", fill = "Gender") +
  ggtitle("Gender-based Salary Distribution by Top Job Titles") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Gender-based Salary Distribution by Top Job Titles



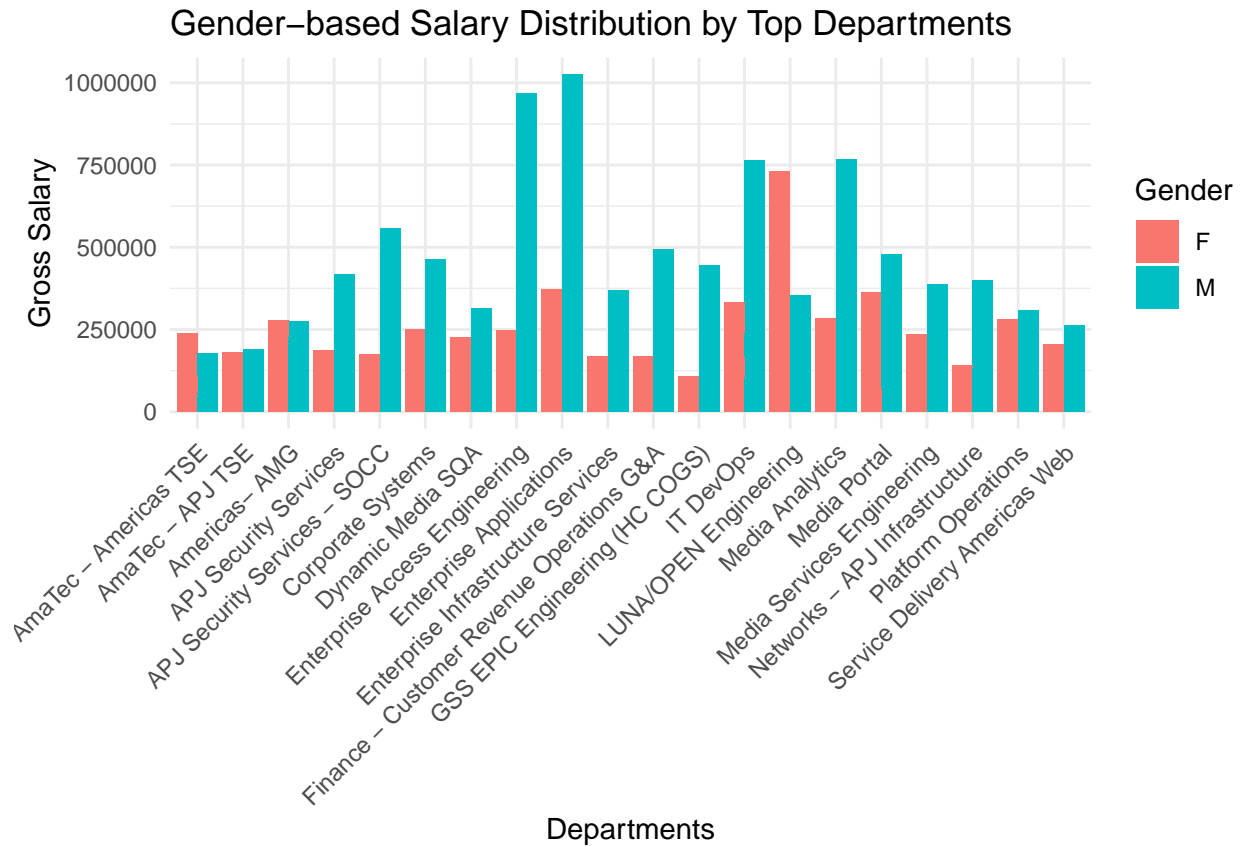
```
#filter by top department
```

```
top_departments <- Employee_monthly_salary %>%
  count(Department) %>%
  top_n(20)
```

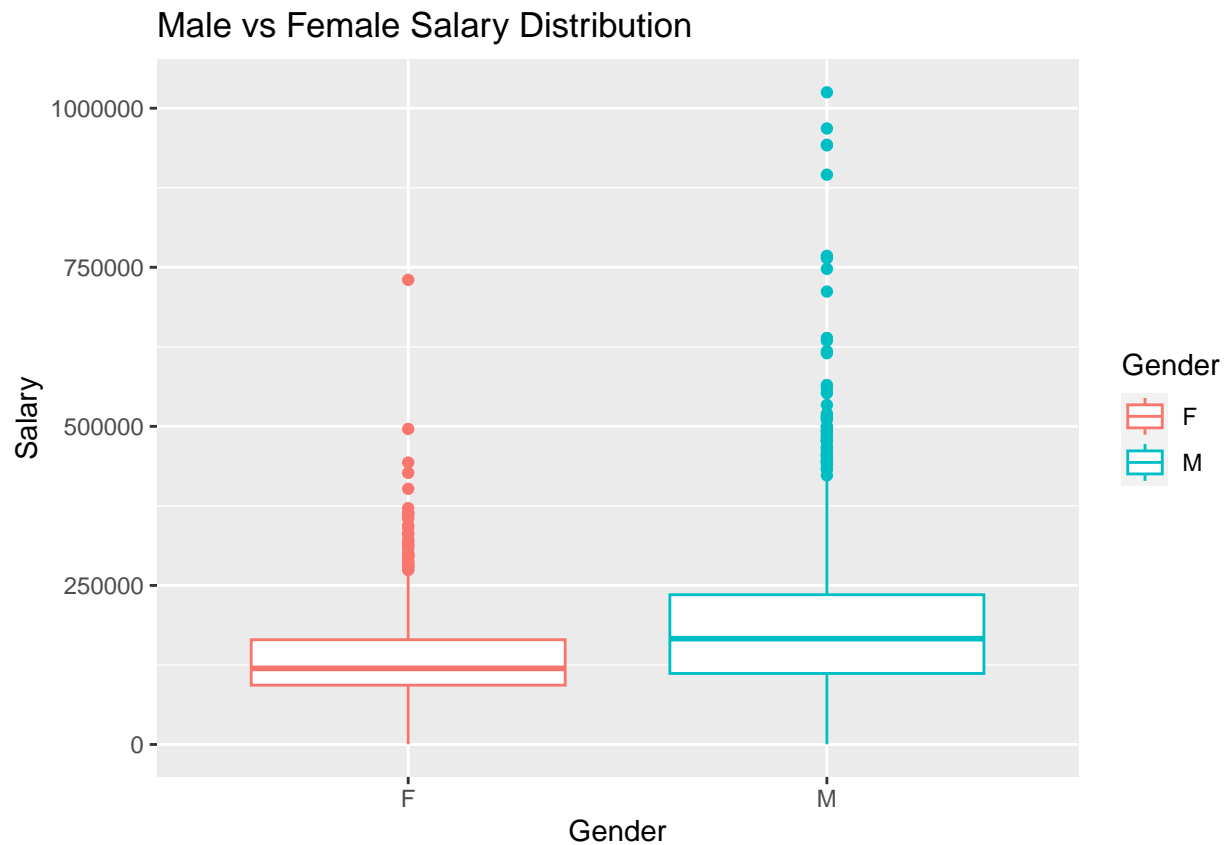
```
## Selecting by n
```

```
#bar chart with filtered job titles
```

```
ggplot(filter(Employee_monthly_salary, Department %in% top_departments$Department),
  aes(x = Department, y = GROSS, fill = Gender)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(x = "Departments", y = "Gross Salary", fill = "Gender") +
  ggtitle("Gender-based Salary Distribution by Top Departments") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



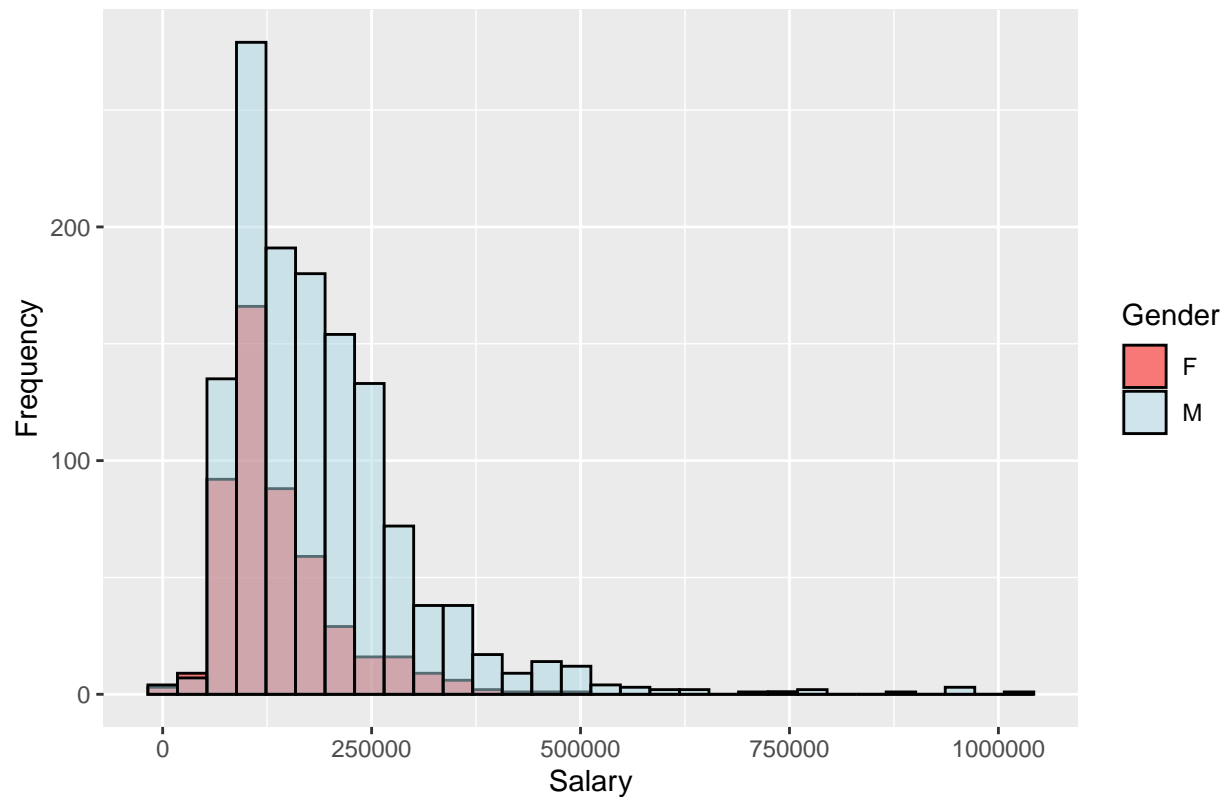
```
# boxplot
ggplot(Employee_monthly_salary, aes(x = Gender, y = GROSS, color = Gender)) +
  geom_boxplot() +
  labs(x = "Gender", y = "Salary", color = "Gender") +
  ggtitle("Male vs Female Salary Distribution")
```



Histogram of Male and Female Salaries

```
#histogram of male and female salaries
ggplot(Employee_monthly_salary, aes(x = GROSS, fill = Gender)) +
  geom_histogram(position = "identity", bins = 30, alpha = 0.5, color = "black") +
  scale_fill_manual(values = c("red", "light blue")) +
  labs(x = "Salary", y = "Frequency", fill = "Gender") +
  ggtitle("Distribution of Salaries by Gender")
```

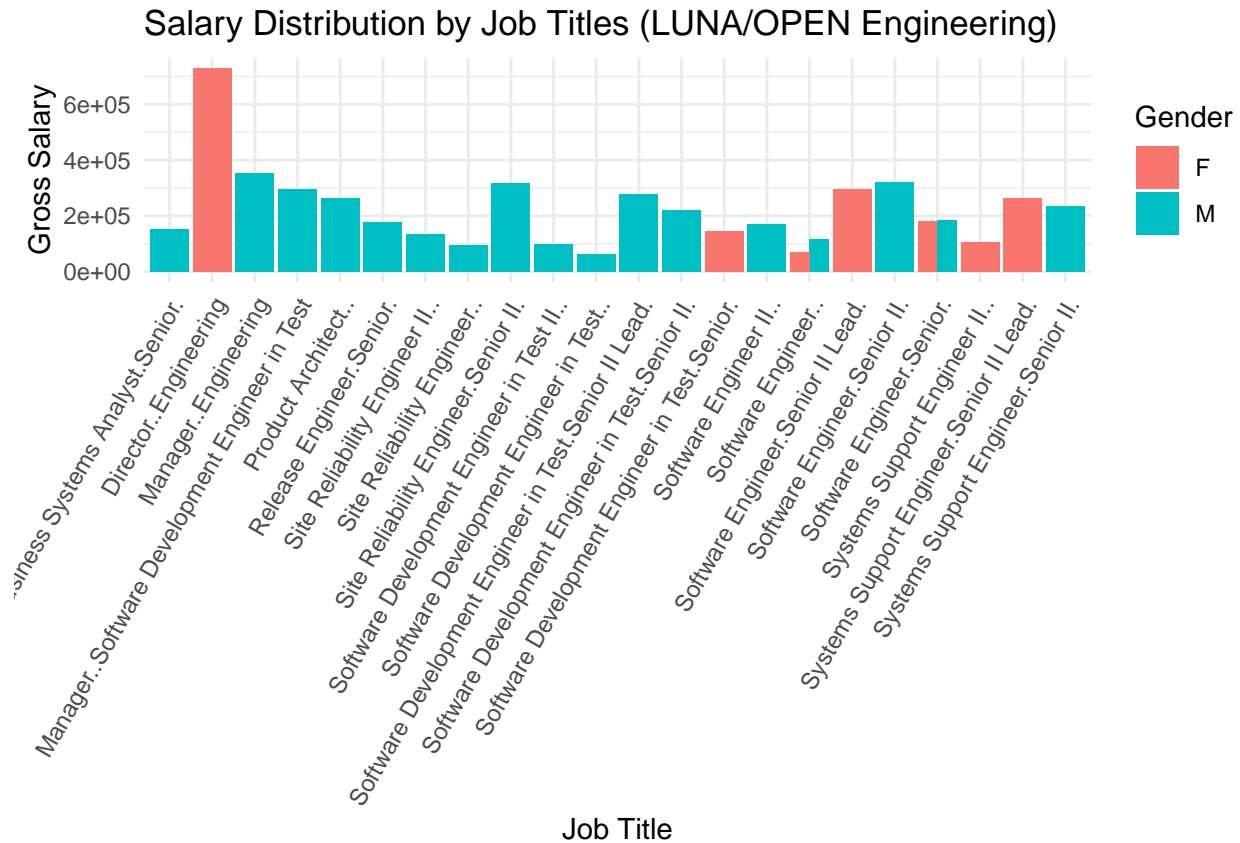
Distribution of Salaries by Gender



Barchart of every Designation in the Department LUNA/OPEN Engineering

```
# Filter dataset by Department
luna_open_data <- Employee_monthly_salary %>%
  filter(Department == "LUNA/OPEN Engineering")

# Create the grouped bar chart
ggplot(luna_open_data, aes(x = Designation, y = GROSS, fill = Gender)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(x = "Job Title", y = "Gross Salary", fill = "Gender") +
  ggtitle("Salary Distribution by Job Titles (LUNA/OPEN Engineering)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 60, hjust = 1))
```

```
# Filter dataset by Department and Job Title
directory_engineers <- Employee_monthly_salary %>%
  filter(Department == "LUNA/OPEN Engineering" & Designation == "Director..Engineering")

# Count the number of Director..Engineering
count_directory_engineers <- nrow(directory_engineers)

# Print the count
print(count_directory_engineers)
```

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

```
# Get the gender of the Director..Engineering
gender <- directory_engineers$Gender

# Print the gender
print(gender)
```

```
## [1] "F"
```

Barchart of every Designation in the Department AmaTec-Americas TSE

```
# Filter dataset by Department
AMA <- Employee_monthly_salary %>%
  filter(Department == "AmaTec - Americas TSE")

# Create the grouped bar chart
ggplot(AMA, aes(x = Designation, y = GROSS, fill = Gender)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(x = "Job Title", y = "Gross Salary", fill = "Gender") +
  ggtitle("Salary Distribution by Job Titles (AmaTec - Americas TSE)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

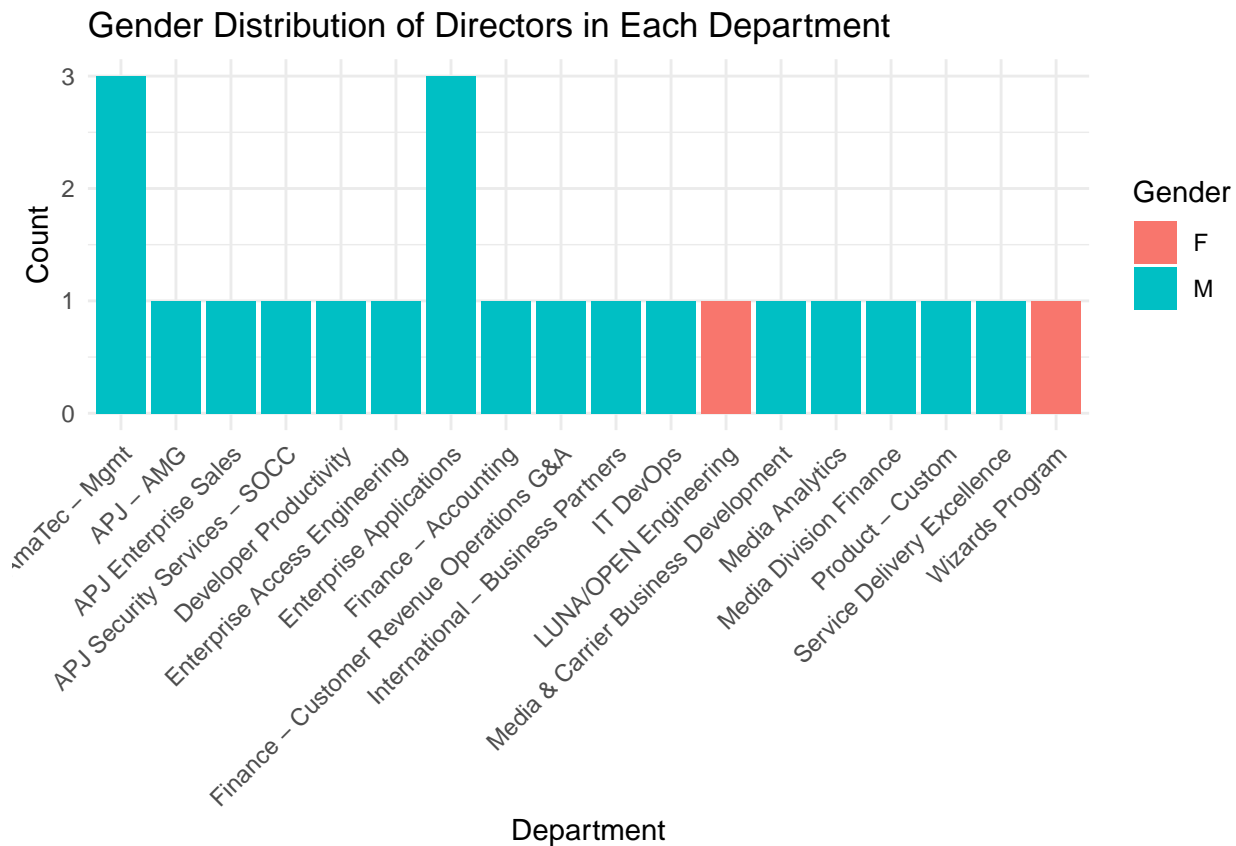


In each department the Directors were graphed to see whether they are male or female

```
# Filter dataset for job titles containing "Director" and count by department and gender
director_counts <- Employee_monthly_salary %>%
  filter(grepl("Director", Designation, ignore.case = TRUE)) %>%
  group_by(Department, Gender) %>%
  summarise(count = n())
```

'summarise()' has grouped output by 'Department'. You can override using the
'.groups' argument.

```
# Create the bar chart
ggplot(director_counts, aes(x = Department, y = count, fill = Gender)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(x = "Department", y = "Count", fill = "Gender") +
  ggtitle("Gender Distribution of Directors in Each Department") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



Look at the wizards program since it contains a female director

```
# Filter dataset by Department
wizard_program <- Employee_monthly_salary %>%
  filter(Department == "Wizards Program")

# Create the grouped bar chart
ggplot(wizard_program, aes(x = Designation, y = GROSS, fill = Gender)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(x = "Job Title", y = "Gross Salary", fill = "Gender") +
  ggtitle("Salary Distribution by Job Titles (Wizards Program)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

