Data 608 - Story 4

Shamecca Marshall

2024-03-19

Objective

I have introduced the term “Data Practitioner” as a generic job descriptor because we have so many different job role titles for individuals whose work activities overlap including Data Scientist, Data Engineer, Data Analyst, Business Analyst, Data Architect, etc. For this story we will answer the question, “How much do we get paid?” Your analysis and data visualizations must address the variation in average salary based on role descriptor and state.

The term “Data Practitioner” encapsulates a broad spectrum of work that bridges data science, requiring both analytical prowess and effective communication skills. It involves translating data into actionable insights and presenting them in a comprehensible manner to diverse audiences, blending elements of both art and science.

Brief

For this project, I examined data sourced from ZipRecruiter throughout 2024. I refined a CSV file to include job roles that align with the realm of a data practitioner. Specifically, I filtered for positions such as Data Analyst, Data Scientist, Business Analyst, and Big Data Engineer.

The findings from this analysis were largely intuitive, yet yielded valuable insights for individuals seeking job opportunities.

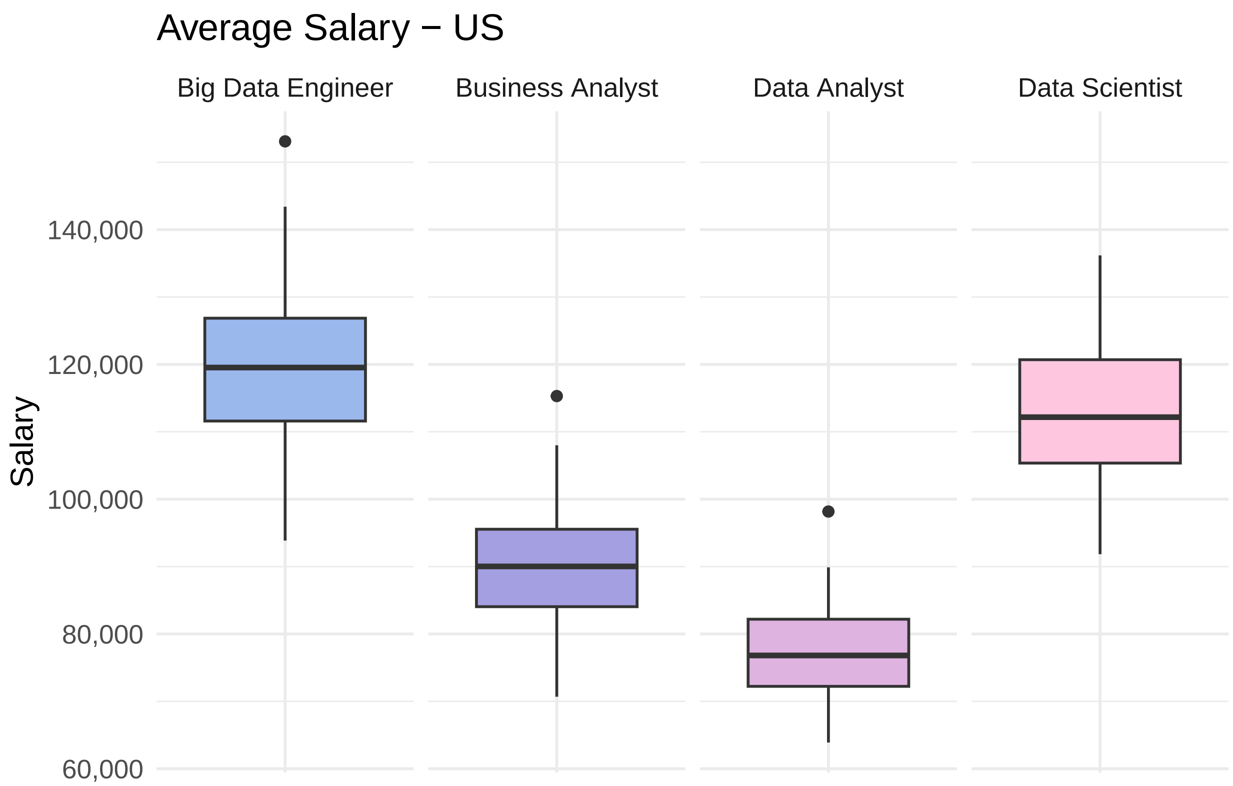
Loading the dataset from the URL

```r
url <- "https://raw.githubusercontent.com/Meccamarshall/Data608/main/Week8/Story4.csv"
data <- read.csv(url)
head(data)
```

```
##        Job_Title State Annual_Salary Monthly.Pay Weekly_Pay Hourly_Wage A_Mean
## 1 Data Scientist    NY        136172       11347       2618          65 112831
## 2 Data Scientist    VT        133828       11152       2573          64 112831
## 3 Data Scientist    CA        131441       10953       2527          63 112831
## 4 Data Scientist    ME        127644       10637       2454          61 112831
## 5 Data Scientist    ID        126275       10522       2428          61 112831
## 6 Data Scientist    WA        125289       10440       2409          60 112831
```

Creating boxplot to view average annual salary by job title

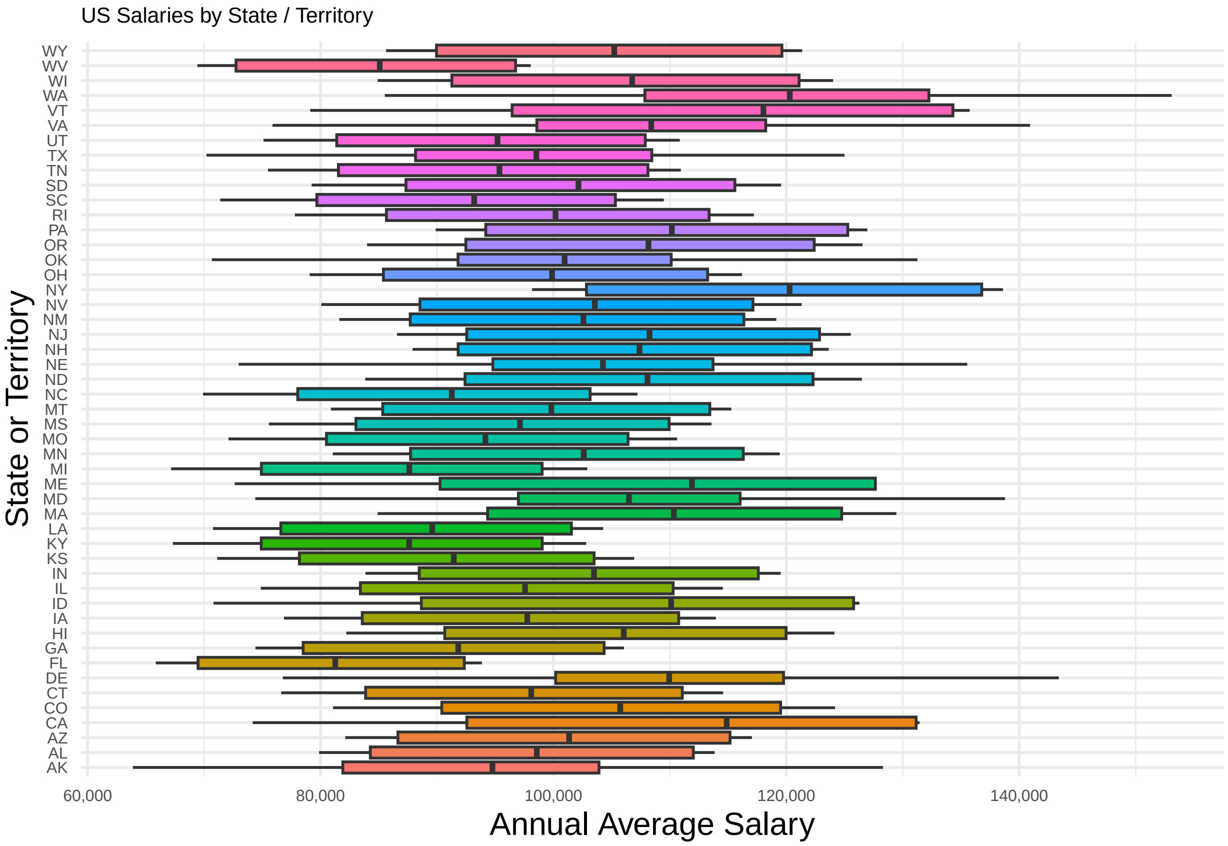
```
palette <- c("#9BB8ED", "#A39FE1", "#DEB3E0", "#FEC6DF")


bp_jobtitle <- ggplot(data, aes(x=" ", y=Annual_Salary, group=Job_Title)) +
  geom_boxplot(aes(fill=Job_Title)) + theme_minimal()
bp_jobtitle <- bp_jobtitle + scale_y_continuous(labels = label_comma())
bp_jobtitle <- bp_jobtitle + facet_grid(. ~ Job_Title)
bp_jobtitle <- bp_jobtitle + scale_fill_manual(values=palette)
bp_jobtitle <- bp_jobtitle + theme(legend.position="none")
bp_jobtitle <- bp_jobtitle + theme(text = element_text(size=12), axis.title=element_text(size=12))
bp_jobtitle <- bp_jobtitle + labs(title = "Average Salary - US", x= " ", y= "Salary")

bp_jobtitle
```



While there isn't a significant variance among the job titles, it's evident that "Big Data Engineer" stands out with notably higher compensation compared to the others. With the state-level data extracted from the dataset and irrelevant state records removed, we can now produce an informative graphic showcasing data across all job titles and states.

Creating boxplot to view average annual salary by state

```
bp_state <- ggplot(data, aes(x=State, y=Annual_Salary, fill=State)) +
  geom_boxplot() + theme_minimal() + coord_flip()
bp_state <- bp_state + scale_y_continuous(labels = label_comma())
bp_state <- bp_state + theme(legend.position="none")
bp_state <- bp_state + theme(text = element_text(size=8), axis.title=element_text(size=12))
bp_state <- bp_state + labs(title = "US Salaries by State / Territory", x= "State or Territory", y= "An
bp_state <- bp_state + theme(plot.title = element_text(size=8))

bp_state
```



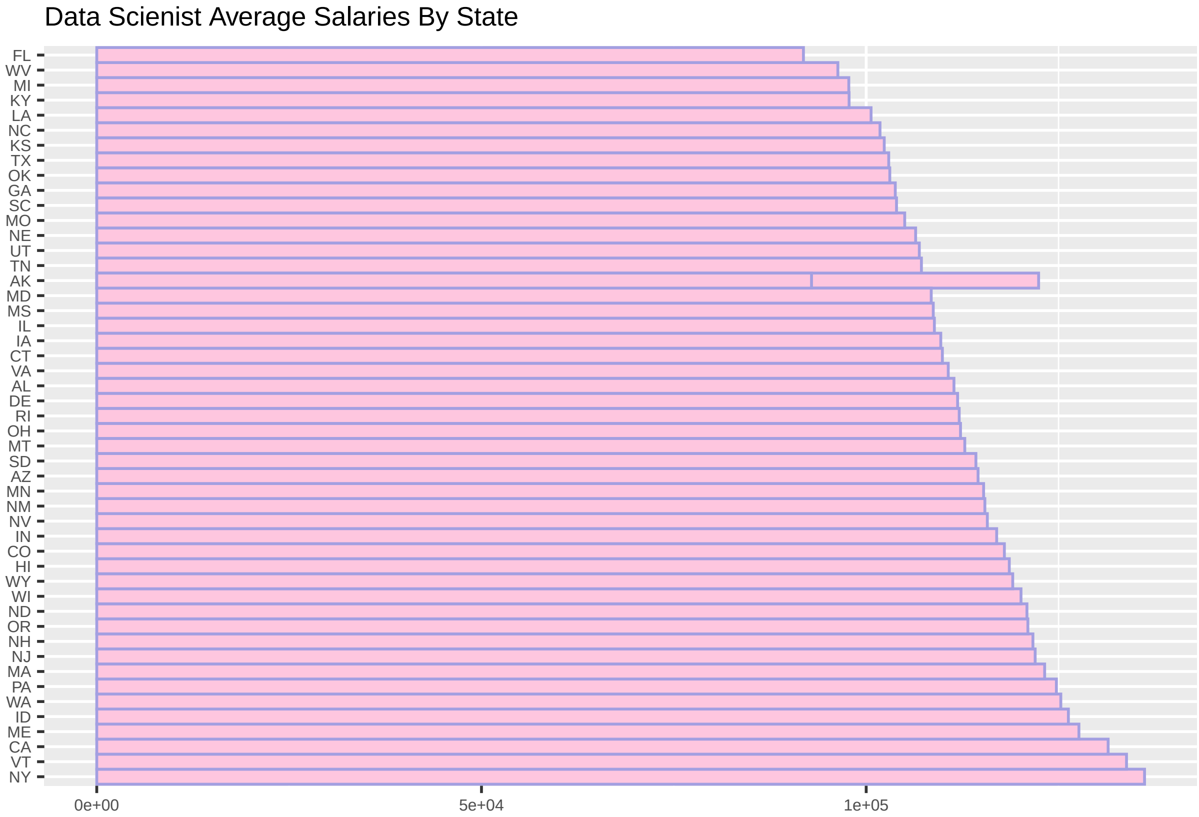
There aren't many unexpected findings here, especially considering the prominence of three leading states that host the country's major tech companies (WA, CA, and NY). However, we would gain valuable insights from a breakdown of each occupation title by state. While box plots are suitable for comparing distributions, they may not be as effective when each occupation is represented by a single figure. Therefore, I've opted for bar charts in the graphics below for better clarity.

Creating new data from filtered data frame.

```
data_ds <- data %>%
  filter(Job_Title == "Data Scientist")

data_da <- data %>%
  filter(Job_Title == "Data Analyst")

data_ba <- data %>%
  filter(Job_Title == "Business Analyst")

data_bde <- data %>%
  filter(Job_Title == "Big Data Engineer")
```

Creating bar graph to view Data Scienist annual salary per state

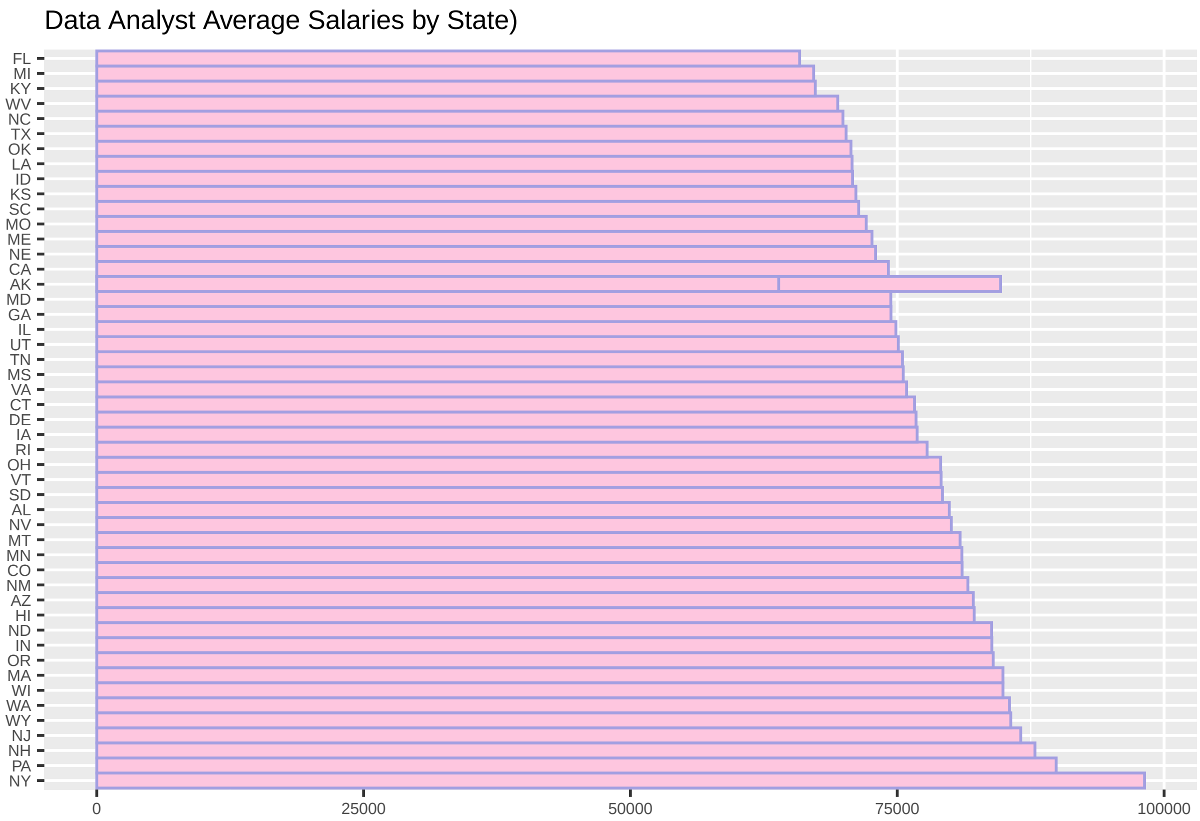
```
ggplot(data_ds) +
  geom_bar(aes(x = reorder(State, -Annual_Salary), y = Annual_Salary, fill = Annual_Salary), stat = "ide
  theme(legend.position = "none", text = element_text(size=8)) +
  labs( title = "Data Scienist Average Salaries By State", x = "", y = "", fill = "Source")
```



The three leading states for Data Scientist salaries are: New York, Vermont, and California. I am a little surprised that DS make more in Vermont than they do in California. This makes me wonder if the demand-supply dynamics for data scientists might differ between the two states. California's tech hub status attracts numerous data science professionals, leading to a more saturated job market and potentially lower average salaries due to higher competition. Conversely, Vermont's smaller tech industry may result in fewer data scientists, thus driving up the average salary due to increased demand. Moreover, regional economic factors, industry concentrations, and state-specific policies regarding incentives or tax structures could contribute to the salary disparity observed between Vermont and California for data scientists.

Creating bar graph to view Data Analyst annual salary per state

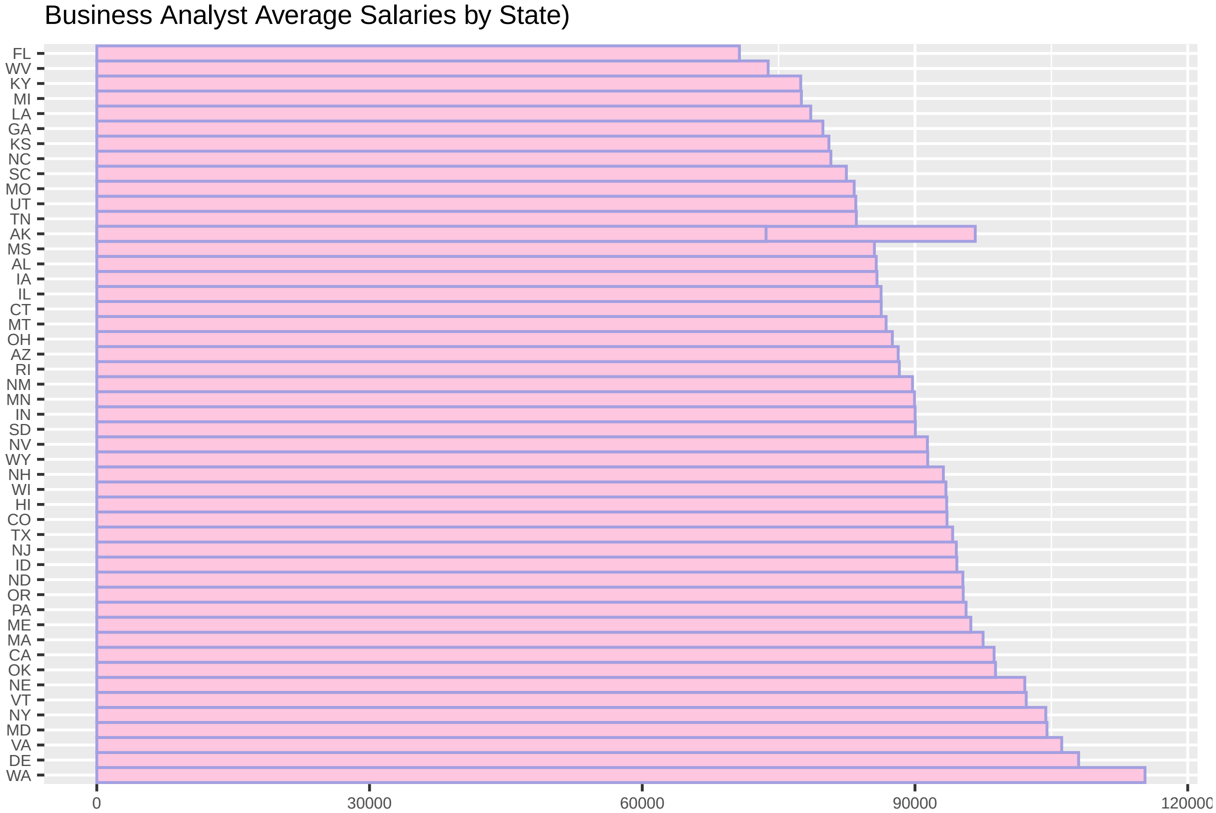
```
ggplot(data_da) +
  geom_bar(aes(x = reorder(State, -Annual_Salary), y = Annual_Salary, fill = Annual_Salary), stat = "ide
  theme(legend.position = "none", text = element_text(size=8)) +
  labs( title = "Data Analyst Average Salaries by State)", x = "", y = "", fill = "Source")
```



The three leading states for Business Analyst salaries are:New York, Pennsylvania, and New Hampshire. These states likely offer the highest annual salaries for data analysts due to a combination of factors. Firstly, New York hosts a thriving financial and tech sector, driving demand for data analysts and consequently offering competitive salaries. Pennsylvania, with its strong presence in healthcare, education, and finance, also provides ample opportunities for data analysts, reflecting in higher pay scales. Additionally, New Hampshire's burgeoning tech industry, coupled with its proximity to major economic hubs like Boston, contributes to elevated salaries for data analysts. These states benefit from robust industries, leading to increased demand for data expertise and thus higher compensation for professionals in this field.

Creating bar graph to view Business Analyst annual salary per state

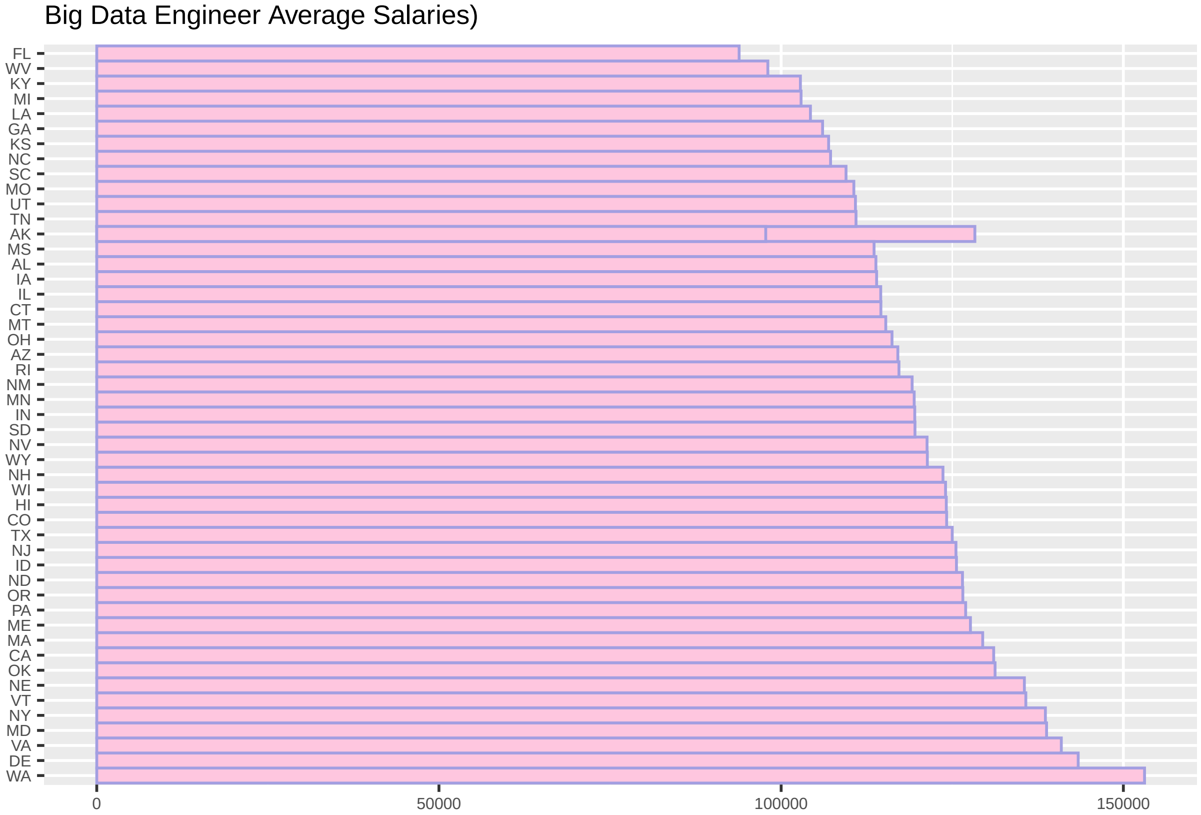
```r
ggplot(data_ba) +
  geom_bar(aes(x = reorder(State, -Annual_Salary), y = Annual_Salary, fill = Annual_Salary), stat = "ide
  theme(legend.position = "none", text = element_text(size=8)) +
  labs( title = "Business Analyst Average Salaries by State)", x = "", y = "", fill = "Source")
```



The three leading states for Business Analyst salaries are: Washington, Delaware, and Maryland. These states often host thriving industries that heavily rely on business analysis, such as technology, finance, and government sectors. In Washington, for instance, the presence of major tech companies like Amazon and Microsoft contributes to a high demand for skilled business analysts. Delaware's status as a financial hub, particularly for banking and corporate sectors, leads to lucrative opportunities for business analysts. Similarly, Maryland, with its concentration of government agencies, biotechnology firms, and defense contractors, offers ample employment prospects for business analysts. Moreover, the cost of living in these states tends to be higher compared to the national average. Employers in these regions often offer competitive salaries to attract and retain talent in the face of elevated living expenses. Additionally, factors such as strong economic growth, favorable business environments, and robust job markets further contribute to the higher salaries observed for business analysts in Washington, Delaware, and Maryland.

Creating bar graph to view Big Data Engineer annual salary per state

```
ggplot(data_bde) +
  geom_bar(aes(x = reorder(State, -Annual_Salary), y = Annual_Salary, fill = Annual_Salary), stat = "ide
  theme(legend.position = "none", text = element_text(size=8)) +
  labs( title = "Big Data Engineer Average Salaries)", x = "", y = "", fill = "Source")
```



The three leading states for Big Data Engineer salaries are: Washington, Delaware, and Virginia. Washington, Delaware, and Virginia emerge as the top-paying states for big data engineers due to a combination of factors. Firstly, these states are home to major technology firms and government agencies, driving demand for professionals skilled in managing and analyzing large datasets. Secondly, their robust economies and high-tech industries often offer competitive compensation packages to attract top talent. Additionally, these states may have a relatively lower cost of living compared to other tech-centric regions like California, allowing companies to allocate more resources towards employee salaries. Lastly, state-specific initiatives, such as tax incentives or investment in technology sectors, could further bolster salaries for big data engineers in Washington, Delaware, and Virginia.

Conclusion

Several key insights can be drawn from the data analysis. Firstly, it's evident that "Big Data Engineer" commands notably higher compensation compared to other job titles within the data science domain. This finding underscores the increasing demand for professionals skilled in managing and analyzing large datasets, particularly in states like Washington, Delaware, and Virginia, where major technology firms and government agencies are prevalent.

Furthermore, the prominence of certain states, such as New York, California, and Washington, in offering competitive salaries across various data-related roles suggests a correlation between regional economic factors and job market dynamics. For instance, the thriving tech sectors in California and Washington drive higher demand for data scientists and big data engineers, resulting in elevated compensation levels. Conversely, the relatively smaller tech industry in Vermont may contribute to higher salaries for data scientists due to increased demand and limited supply, despite the state's lower cost of living compared to tech-centric regions.

Additionally, the breakdown of top-paying states for specific job titles, such as New York for data scientists and Delaware for business analysts, reflects the influence of industry concentrations, economic growth, and state-specific policies on salary disparities. States with thriving industries related to finance, technology, and government tend to offer higher compensation to attract and retain talent.

Overall, the analysis highlights the complex interplay between regional factors, industry demand, and job market dynamics in shaping salary trends for data-related roles across different states. By understanding these nuances, employers and job seekers can make more informed decisions regarding talent acquisition and career opportunities within the data science domain.