Project Report

Pletiuk R M

2023-11-12

Project "CodeCurrencies: Decoding IT Salaries through R Analysis"

In this project, I aim to analyze the salaries of IT professionals using R software. I will compile and analyze a dataset, visualizing various aspects of salaries through diverse graphs. Additionally, I will calculate descriptive statistics, construct at least 7 simple linear regression models, and build 5 multiple linear regression models to explore relationships between economic variables and IT salaries. Finally, I will create a report in R, presenting the findings and drawing conclusions on the factors influencing IT salaries.

Plan:

- Descriptive statistics
- Plots
- Pair regression
- Multiple regression

Descriptive statistics (part 1)

Descriptive statistics of variables:

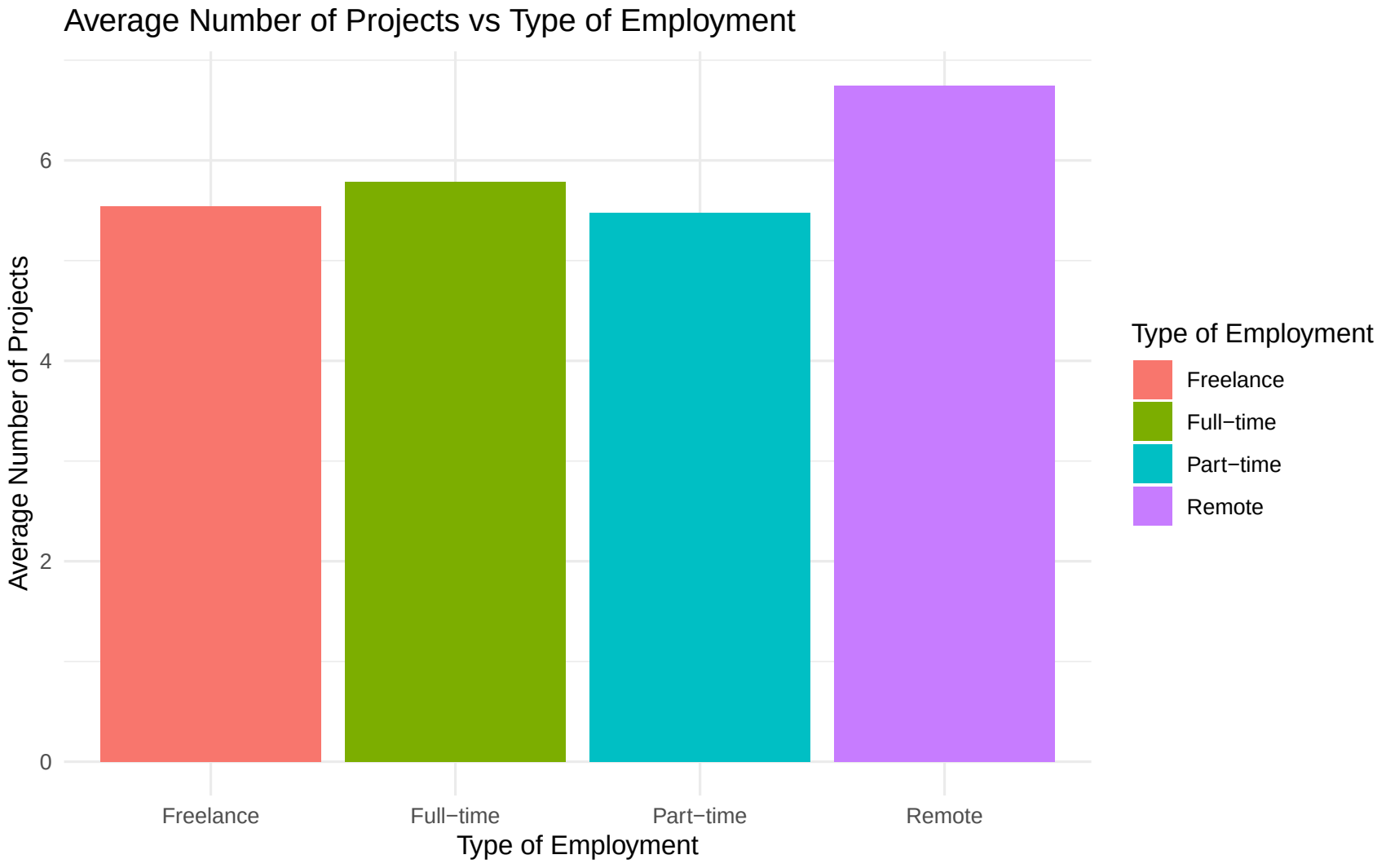
```
##      Salary            Age        Number_of_projects
##  Min.   :12614   Min.   :19.00   Min.   : 1.000
##  1st Qu.:33876   1st Qu.:24.00   1st Qu.: 4.000
##  Median :44534   Median :26.00   Median : 5.000
##  Mean   :45740   Mean   :26.39   Mean   : 5.717
##  3rd Qu.:54822   3rd Qu.:28.00   3rd Qu.: 7.000
##  Max.   :85722   Max.   :36.00   Max.   :11.000
```

Descriptive statistics (part 2)

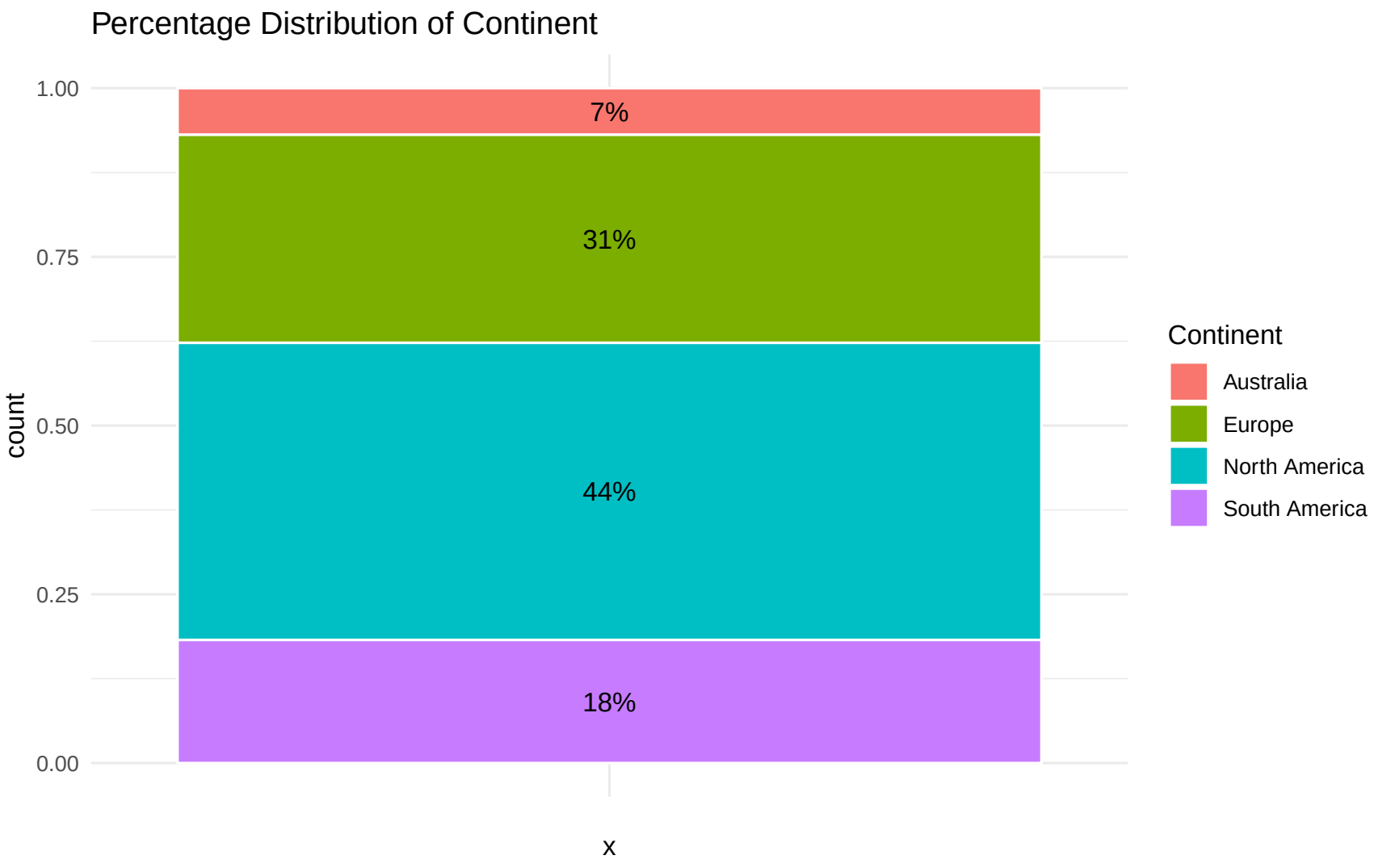
Descriptive statistics of variables:

```
##    Experience      Hour_on_week    Job_satisfaction
##  Min.   :1.000   Min.   :20.00   Min.   :2.500
##  1st Qu.:3.000   1st Qu.:35.00   1st Qu.:3.400
##  Median :5.000   Median :40.00   Median :3.900
##  Mean   :4.409   Mean   :37.16   Mean   :3.958
##  3rd Qu.:6.000   3rd Qu.:40.00   3rd Qu.:4.550
##  Max.   :8.000   Max.   :45.00   Max.   :5.000
```

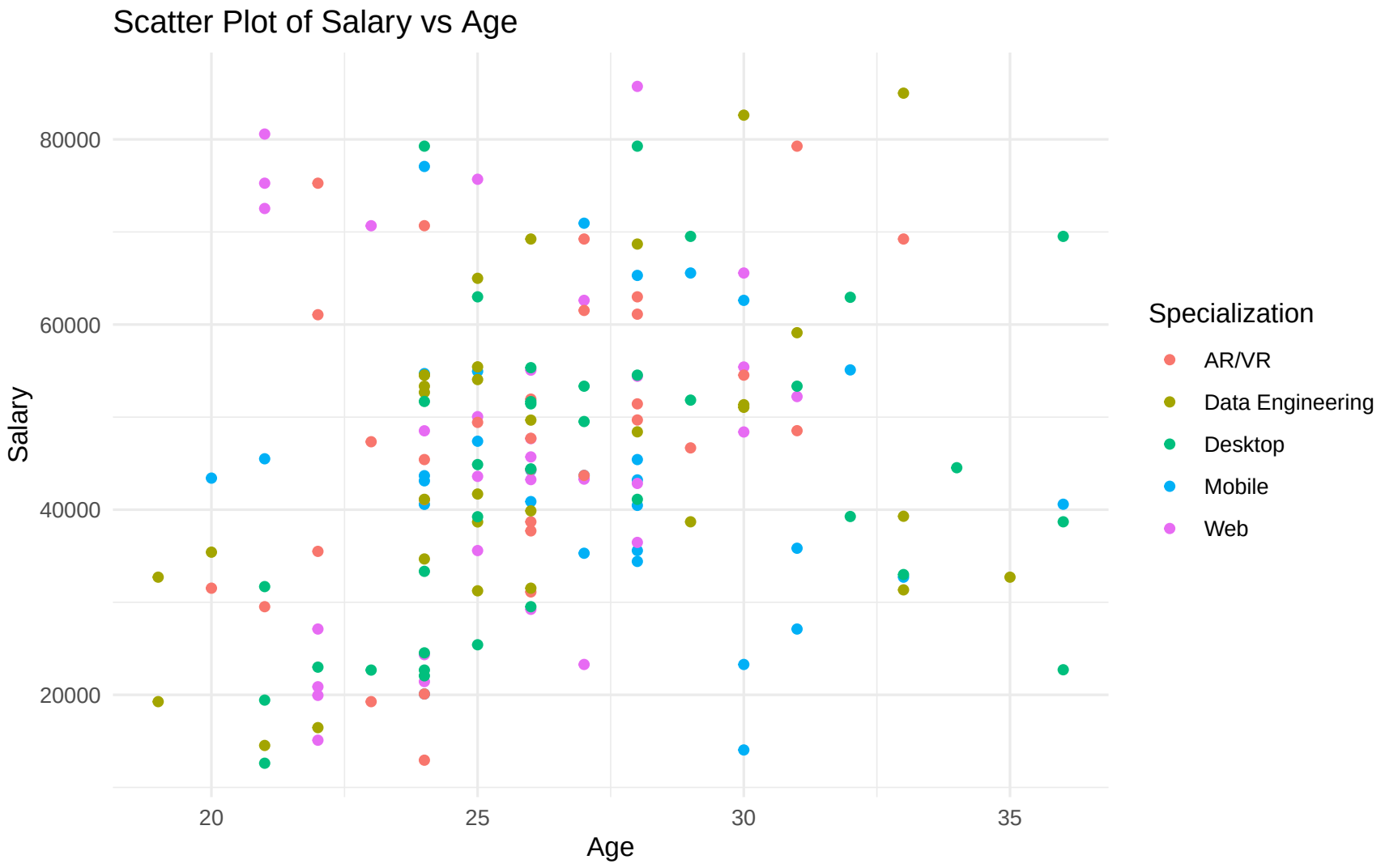
Plot “Average Number of Projects vs Type of Employment”



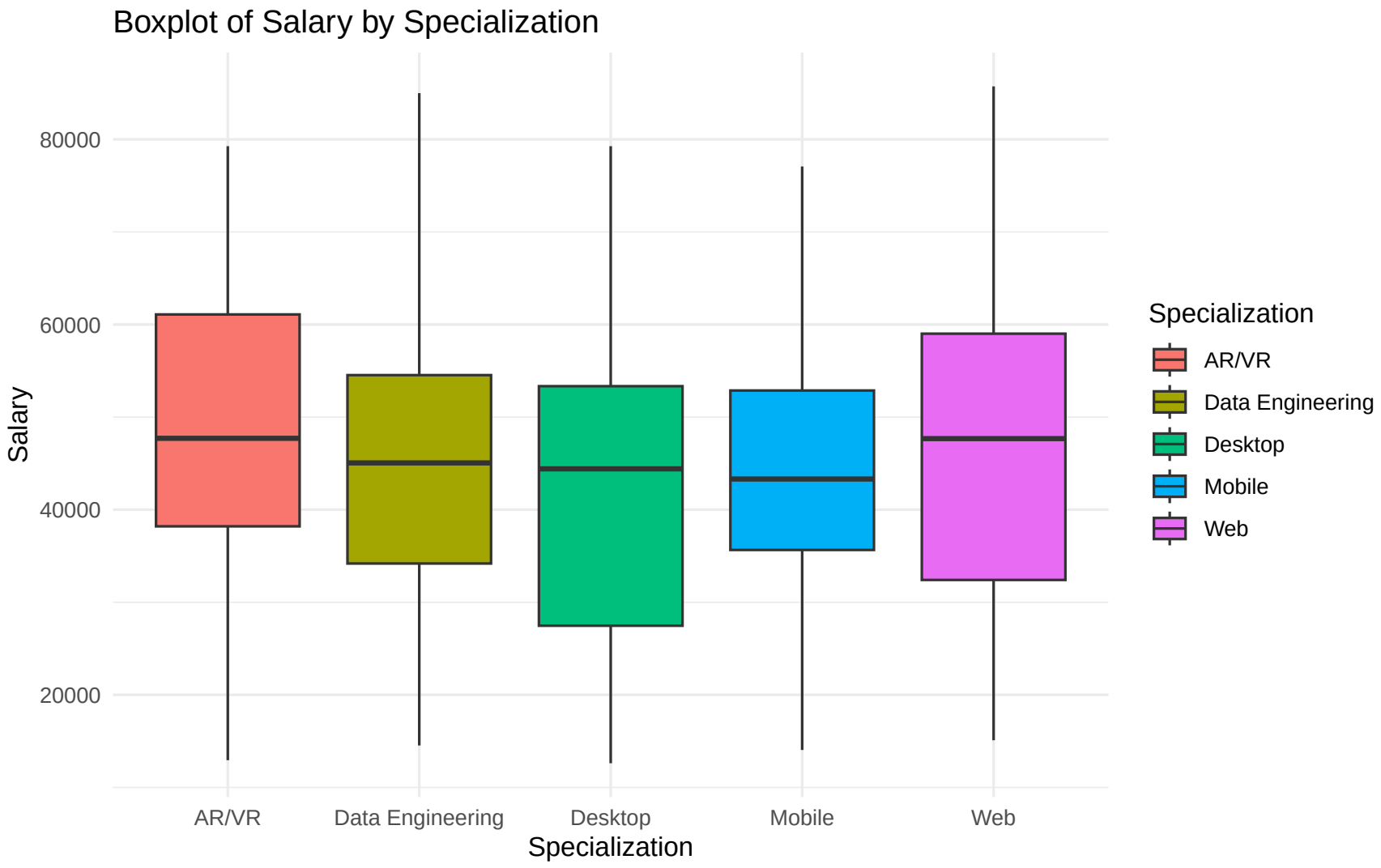
Plot “Percentage Distribution of Continent”



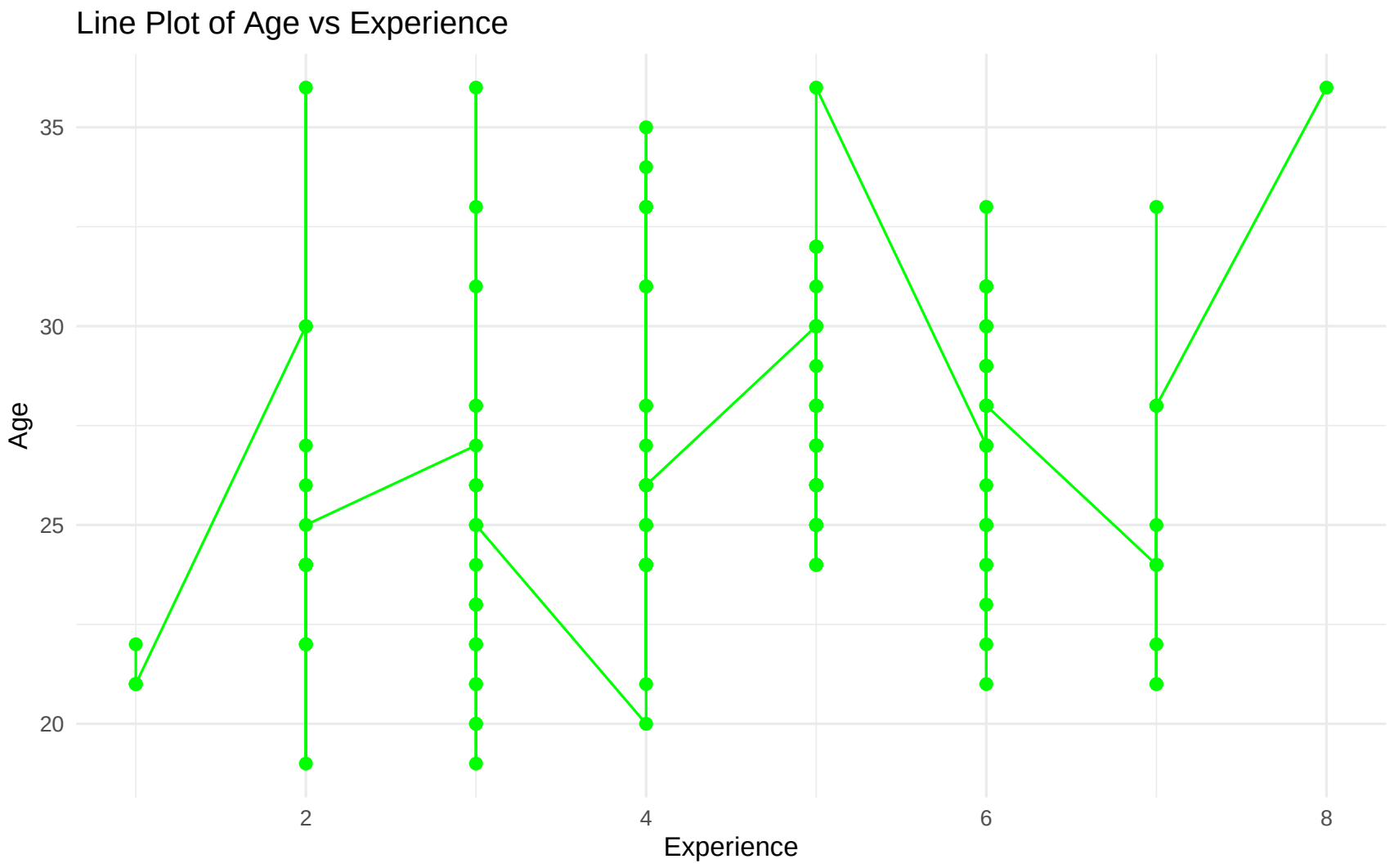
Plot “Scatter Plot of Salary vs Age”



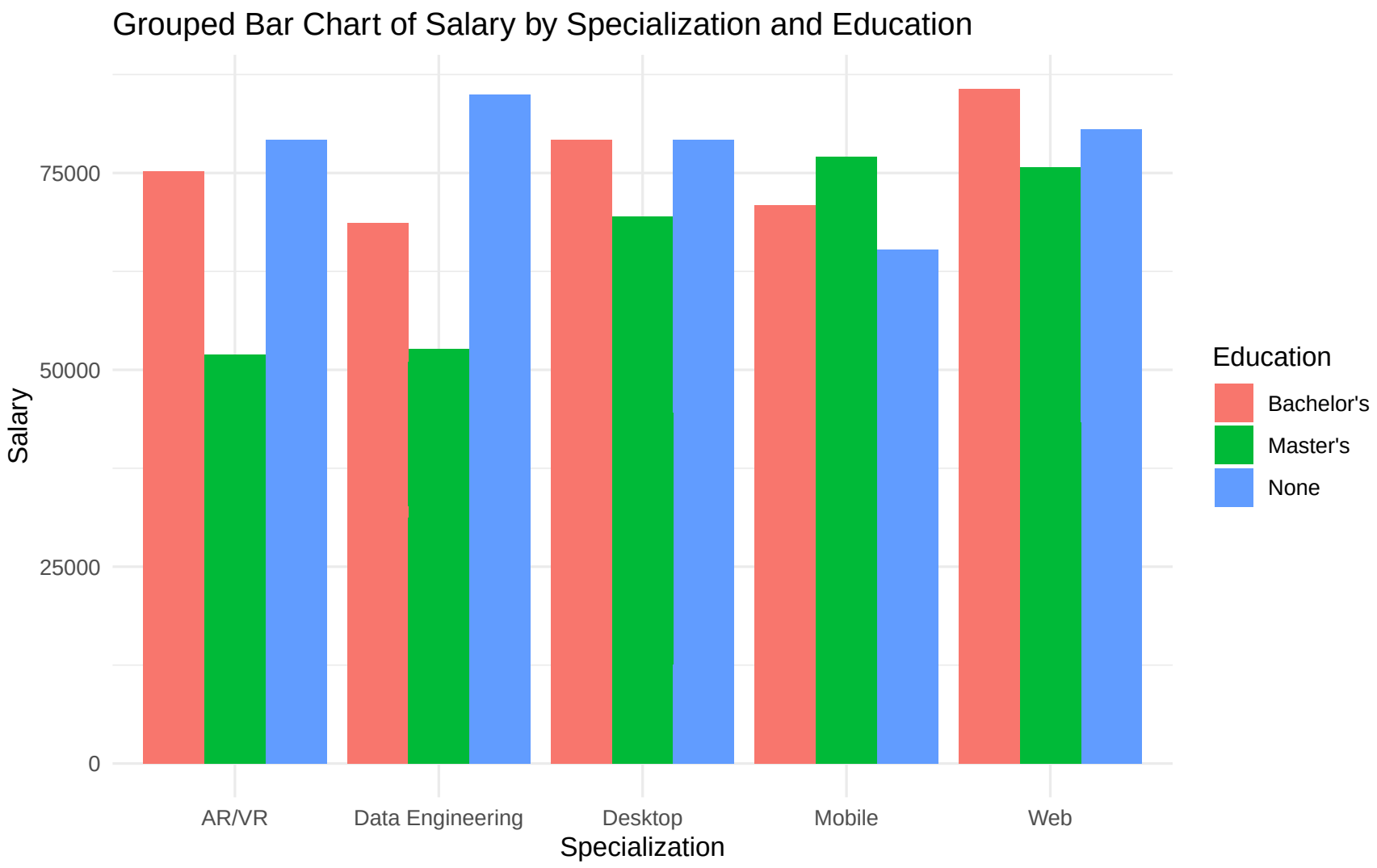
Plot “Boxplot of Salary by Specialization”



Plot “Line Plot of Age vs Experience”



Plot “Salary by Specialization and Education”



These are the pairwise regressions

1. Salary $\sim$ Experience
2. Salary $\sim$ Hour on week
3. Salary $\sim$ Age
4. Salary $\sim$ Number of projects
5. Salary $\sim$ Job satisfaction
6. Salary $\sim$ Additional profit
7. Number of projects $\sim$ Experience

Linear pair Regression: Experience vs Salary (1)

- Plot Linear pair Regression

Linear pair Regression: Experience vs Salary (2)

Call: lm(formula = Salary ~ Experience, data = my_data3)

Residuals: Min 1Q Median 3Q Max -22483 -6146 -44 3900 27790

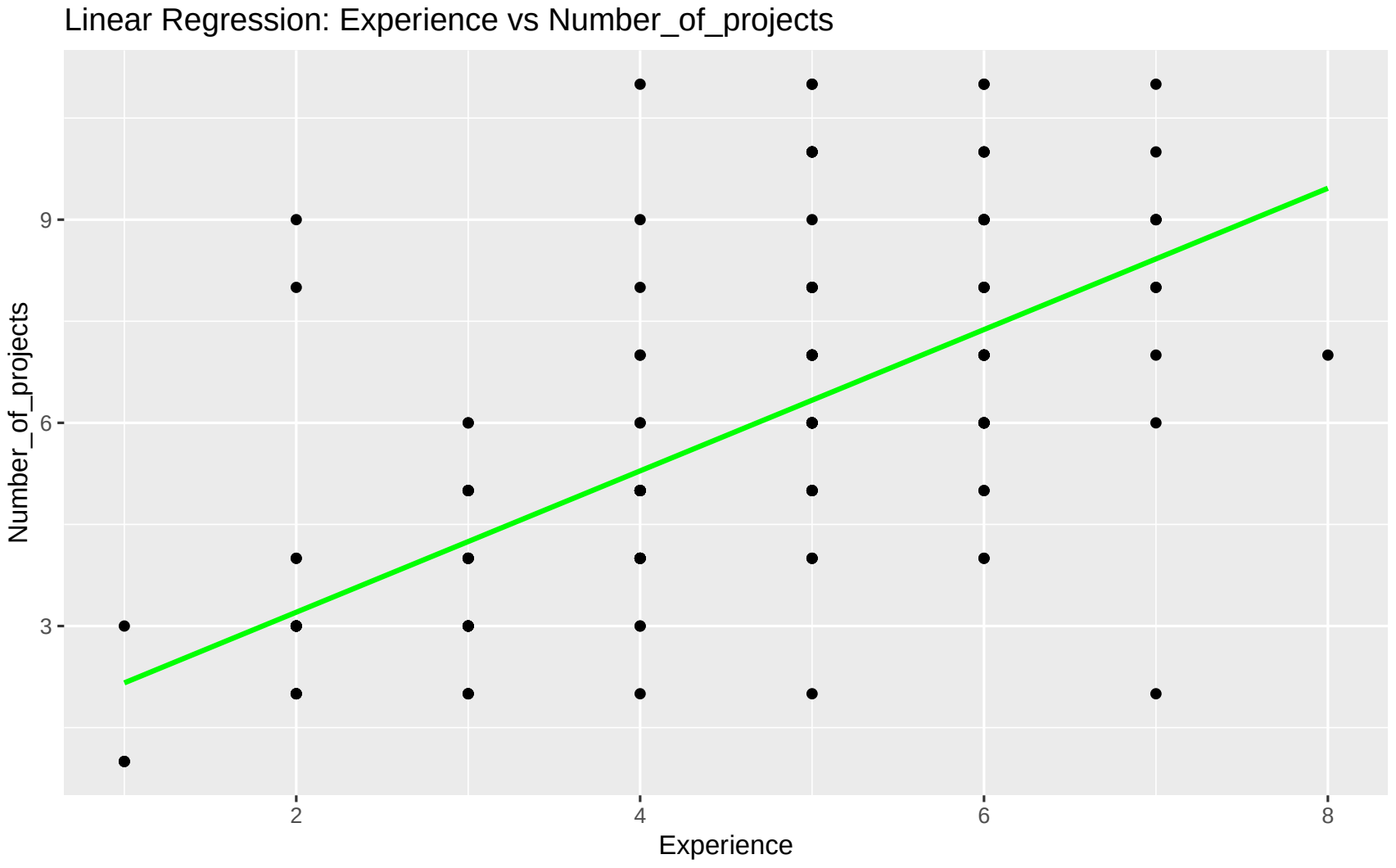
Coefficients: Estimate Std. Error t value Pr(>|t|)
(Intercept) 2974 2034 1.462 0.146
Experience 9700 436 22.246 <2e-16 *** — Signif. codes: 0 ‘’ ***0.001*** ’’ *0.01* ” 0.05 ‘.’ 0.1 ’ ’ 1

Residual standard error: 8370 on 157 degrees of freedom Multiple R-squared: 0.7592, Adjusted R-squared: 0.7576 F-statistic: 494.9 on 1 and 157 DF, p-value: < 2.2e-16

Regression Equation: Salary = 2973.496 + 9700.305 * Experience

Linear pair Regression: Number of projects vs Experience (1)

- Plot Linear pair Regression



Linear pair Regression: Number of projects vs Experience (2)

Call: lm(formula = Number_of_projects ~ Experience, data = my_data3)

Residuals: Min 1Q Median 3Q Max -6.4205 -1.2688 -0.2905 0.7529 5.7962

Coefficients: Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.11709 0.45778 2.44 0.0158 *
Experience 1.04334 0.09815 10.63 <2e-16 *** — Signif. codes: 0 '*** **0.001** ** *0.01* * 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.884 on 157 degrees of freedom Multiple R-squared: 0.4185, Adjusted R-squared: 0.4148 F-statistic: 113 on 1 and 157 DF, p-value: < 2.2e-16

Regression Equation: Number_of_projects = 1.117088 + 1.043342 * Experience

Other pairwise regressions (1)

- Salary ~ Hour on week:

```
##                  Estimate Std. Error  t value   Pr(>|t|)
## (Intercept)  24365.9714 11210.5329 2.173489 0.03124202
## Hour_on_week   575.2387   299.5522 1.920329 0.05662923
```

Regression Equation: Salary vs Hour on week: Salary = 24365.97 + 575.2387 * Hour_on_week

- Salary ~ Age:

```
##                 Estimate Std. Error  t value    Pr(>|t|)
## (Intercept) 19462.8835   9633.995 2.020230 0.045059852
## Age           995.7343    361.615 2.753575 0.006590452
```

Regression Equation: Salary vs Age: Salary = 19462.88 + 995.7343 * Age

Other pairwise regressions (2)

- Salary ~ Number of projects

```
##                      Estimate Std. Error   t value     Pr
## (Intercept)         19383.729  2551.3338  7.597488 2.5542
## Number_of_projects   4610.216   410.0642 11.242670 6.9233
```

Regression Equation: Salary vs Number of projects: Salary = 19383.73 + 4610.216 * Number_of_projects

- Salary ~ Job satisfaction

```
##                       Estimate Std. Error   t value     Pr(
## (Intercept)       43349.9835   7838.245 5.5305724 1.3054
## Job_satisfaction    603.9282   1950.736 0.3095899 7.5728
```

Regression Equation: Salary vs Job satisfaction: Salary = 43349.98 + 603.9282 * Job_satisfaction

Other pairwise regressions (3)

- Salary ~ Additional profit

```
##                          Estimate  Std. Error   t value
## (Intercept)          32841.68327 2311.176973 14.209939 5.45
## Additional_profit       19.72827    3.021613  6.529053 8.70
```

Regression Equation: >Salary vs Additional profit: Salary = 32841.68 + 19.72827 * Additional_profit

Multiple regressions

- Salary $\sim$ Experience + Hour_on_week
- Salary $\sim$ Experience + Number_of_projects
- Salary $\sim$ Experience + Additional_profit
- Salary $\sim$ Experience + Job_satisfaction
- Salary $\sim$ Hour_on_week + Number_of_projects + Additional_profit + Job_satisfaction

Multiple Regression: Experience + Number_of_projects vs Salary (1)

Call: lm(formula = Salary ~ Experience + Number_of_projects, data = my_data)

Residuals: Min 1Q Median 3Q Max -22064.2 -6044.8 -354.3 4736.1 25013.1

Coefficients: Estimate Std. Error t value Pr(>|t|)
(Intercept) 1592.1 1996.5 0.797 0.426394
Experience 8410.1 551.0 15.264 < 2e-16 ***Number_of_projects***
1236.6 341.6 3.620 0.000398 — Signif. codes: 0 '***' 0.001 '**' *0.01* '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8064 on 156 degrees of freedom Multiple R-squared: 0.7778, Adjusted R-squared: 0.775 F-statistic: 273.1 on 2 and 156 DF, p-value: < 2.2e-16

Regression Equation: Salary = 1592.101 + 8410.104 * Experience + 1236.603 * Number_of_projects

Other Multiple Regressions (1)

- Salary ~ Experience + Hour_on_week:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-9184.8841	5694.7392	-1.612872	1.087935e-01
Experience	9628.5160	431.4620	22.316022	2.048344e-50
Hour_on_week	335.7323	147.1601	2.281409	2.387699e-02

Regression Equation: Salary = -9184.884 + 8410.104 * Experience + 1236.603 * Hour_on_week

- Salary ~ Experience + Additional_profit

	Estimate	Std. Error	t value	Pr(>|t|
(Intercept)	2608.81644	2035.195989	1.281850	2.017970e-01
Experience	9326.76970	490.078612	19.031171	1.641173e-42
Additional_profit	3.07661	1.879314	1.637092	1.036270e-01

Regression Equation: Salary = 2608.816 + 8410.104 * Experience + 1236.603 * Additional_profit

Other Multiple Regressions (2)

- Salary $\sim$ Experience $+$ Job_satisfaction

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5846.8343	4205.9455	1.3901355	1.664680e-01
Experience	9721.9655	437.4611	22.2236112	3.354983e-50
Job_satisfaction	-750.1113	960.7357	-0.7807676	4.361223e-01

Regression Equation: Salary $=$ 2608.816 $+$ 8410.104 * Experience $+$ 1236.603 * Job_satisfaction

Other Multiple Regressions (3)

- Salary ~ Hour_on_week + Number_of_projects + Additional_profit + Job_satisfaction

```
##                          Estimate   Std. Error    t value
## (Intercept)          8415.082766 10240.430647  0.8217509 4
## Hour_on_week          385.455932   226.103710  1.7047749 9
## Number_of_projects   4314.487916   526.189921  8.1994879 8
## Additional_profit       2.618995     3.271784  0.8004794 4
## Job_satisfaction    -852.855867  1455.897523 -0.5857939 5
```

Regression Equation: Salary = 8415.083 + 385.4559 * Hour_on_week + 4314.488 * Number_of_projects + 2.618995 * Additional_profit + -852.8559 * Job_satisfaction

Conclusion

Insights and Trends

The analysis revealed key insights into IT salaries, showcasing trends through descriptive statistics and diverse visualizations.

Regression Models

Constructed simple and multiple linear regression models provided a deeper understanding of the impact of economic variables on IT salaries, offering valuable predictive capabilities.

Future Considerations

These findings have implications for strategic HR decisions, and future research could explore industry-specific nuances for a more nuanced perspective.