CASEHW1

2024-10-11

install.packages("ggplot2") install.packages("dplyr")

```
# Load necessary libraries
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
getwd()
```

```
## [1] "/Users/avikalpkarrahe/Desktop/UCD 24-25/Fall 24/Triple I/CASE 1 HW"
```

```
# Load the dataset (adjust the path as needed)
load("casedata.Rdata")

# Check the structure and summary of the data
str(Expedia)
```

```
## 'data.frame':    25000 obs. of  5 variables:
##  $ PricePerNight: num  253 241 272 241 264 259 260 261 250 265 ...
##  $ Region       : chr  "Las Vegas" "Hawaii" "Miami" "Washinton DC" ...
##  $ UserIncome   : num  32000 49000 28000 107000 54000 28000 58000 67000 61000 3800
0 ...
##  $ Booked?      : num  0 0 0 1 1 0 1 1 0 1 ...
##  $ Nights       : num  0 0 0 2 3 0 4 3 0 4 ...
```

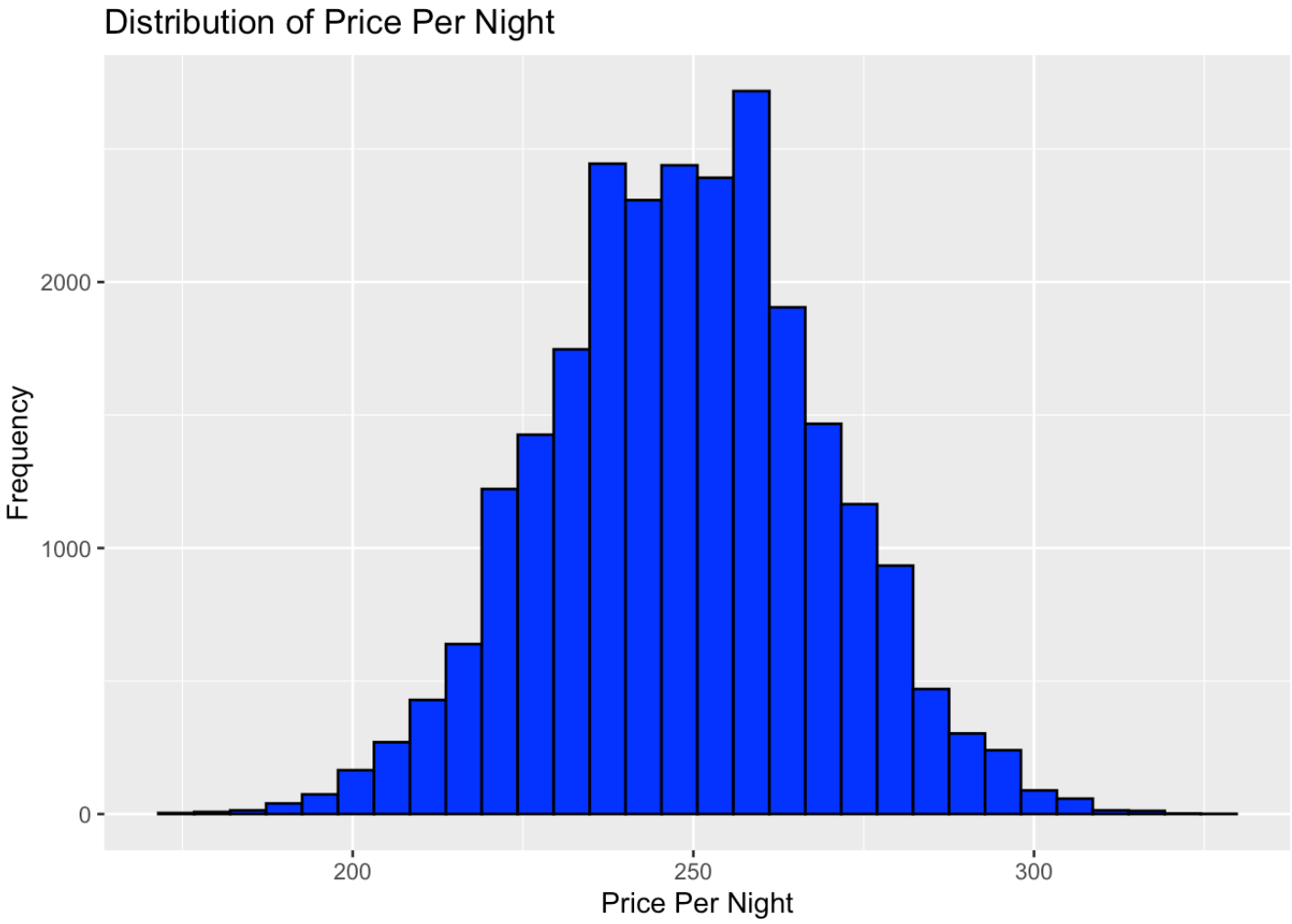
```
summary(Expedia)
```

```
##  PricePerNight      Region            UserIncome        Booked?
##  Min.   :172.0   Length:25000       Min.   :  4000   Min.   :0.0000
##  1st Qu.:235.0   Class :character   1st Qu.: 31000   1st Qu.:0.0000
##  Median :249.0   Mode  :character   Median : 45000   Median :0.0000
##  Mean   :248.9                      Mean   : 52040   Mean   :0.2598
##  3rd Qu.:263.0                      3rd Qu.: 65000   3rd Qu.:1.0000
##  Max.   :325.0                      Max.   :363000   Max.   :1.0000
##      Nights
##  Min.   :0.0000
##  1st Qu.:0.0000
##  Median :0.0000
##  Mean   :0.7672
##  3rd Qu.:2.0000
##  Max.   :6.0000
```

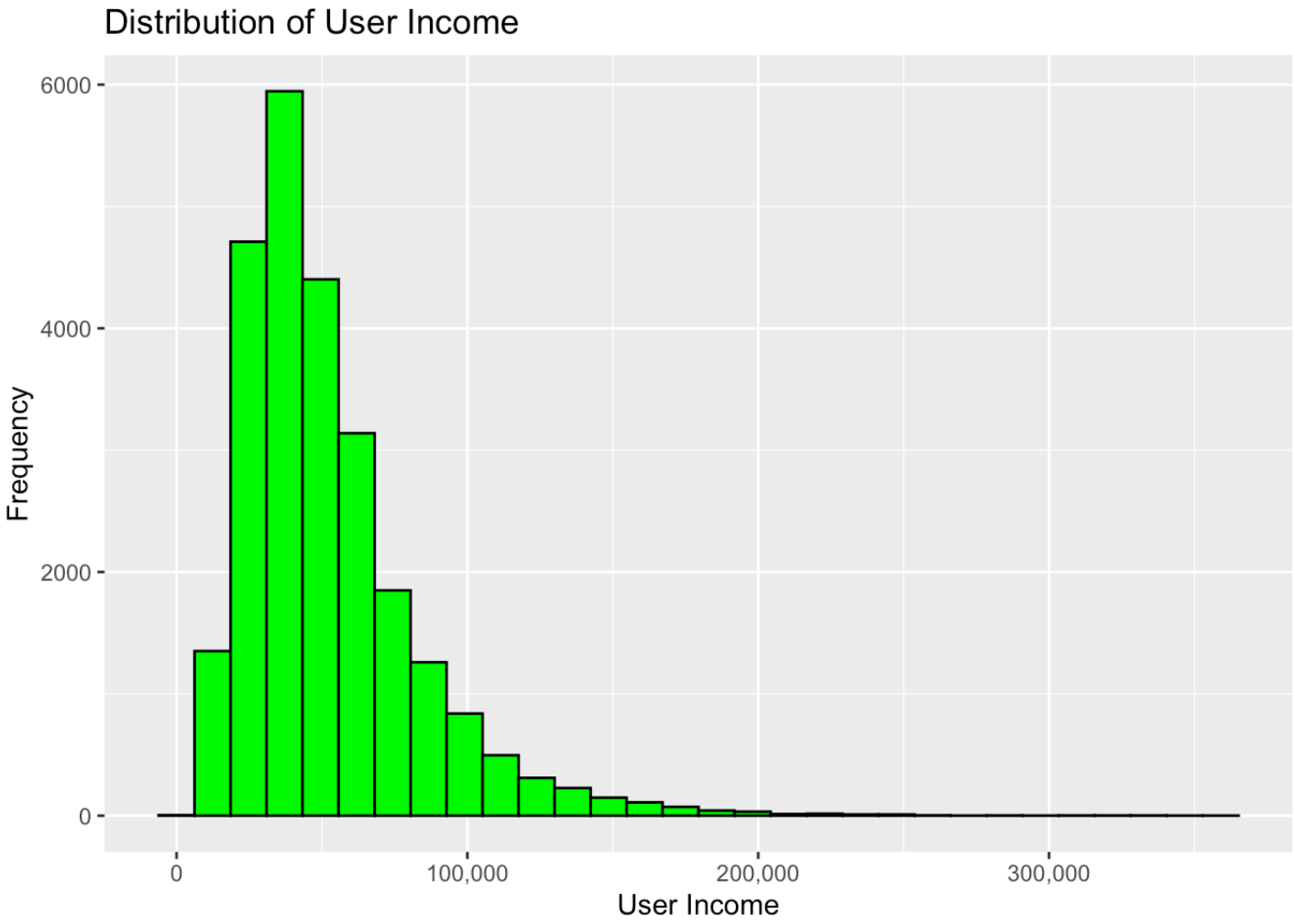
```
# Check for missing values
colSums(is.na(Expedia))
```

```
## PricePerNight        Region    UserIncome       Booked?        Nights
##             0             0             0             0             0
```

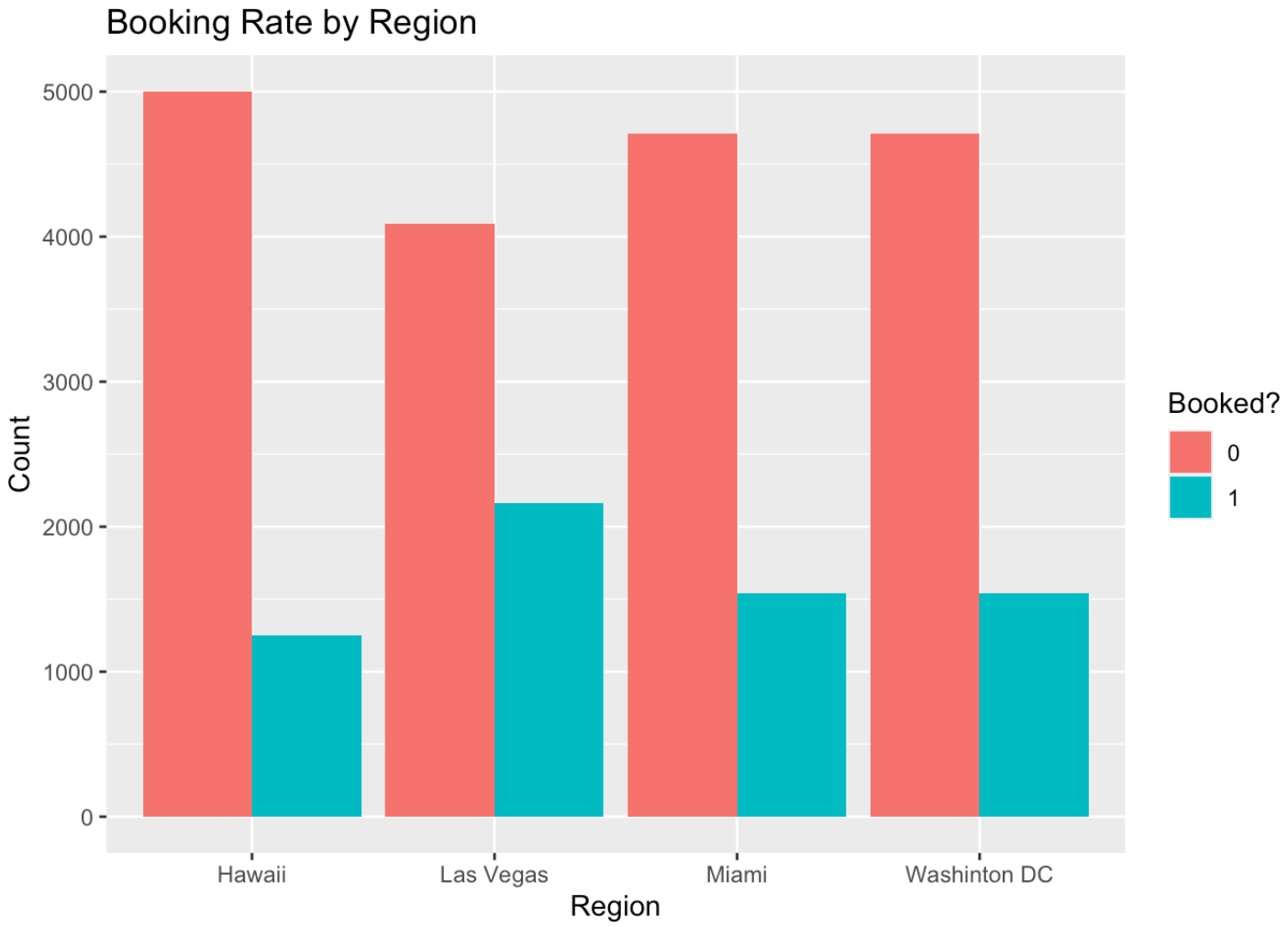
```
# Histogram for Price Per Night
ggplot(Expedia, aes(x = PricePerNight)) +
  geom_histogram(bins = 30, fill = "blue", color = "black") +
  labs(title = "Distribution of Price Per Night", x = "Price Per Night", y = "Frequen
cy")
```



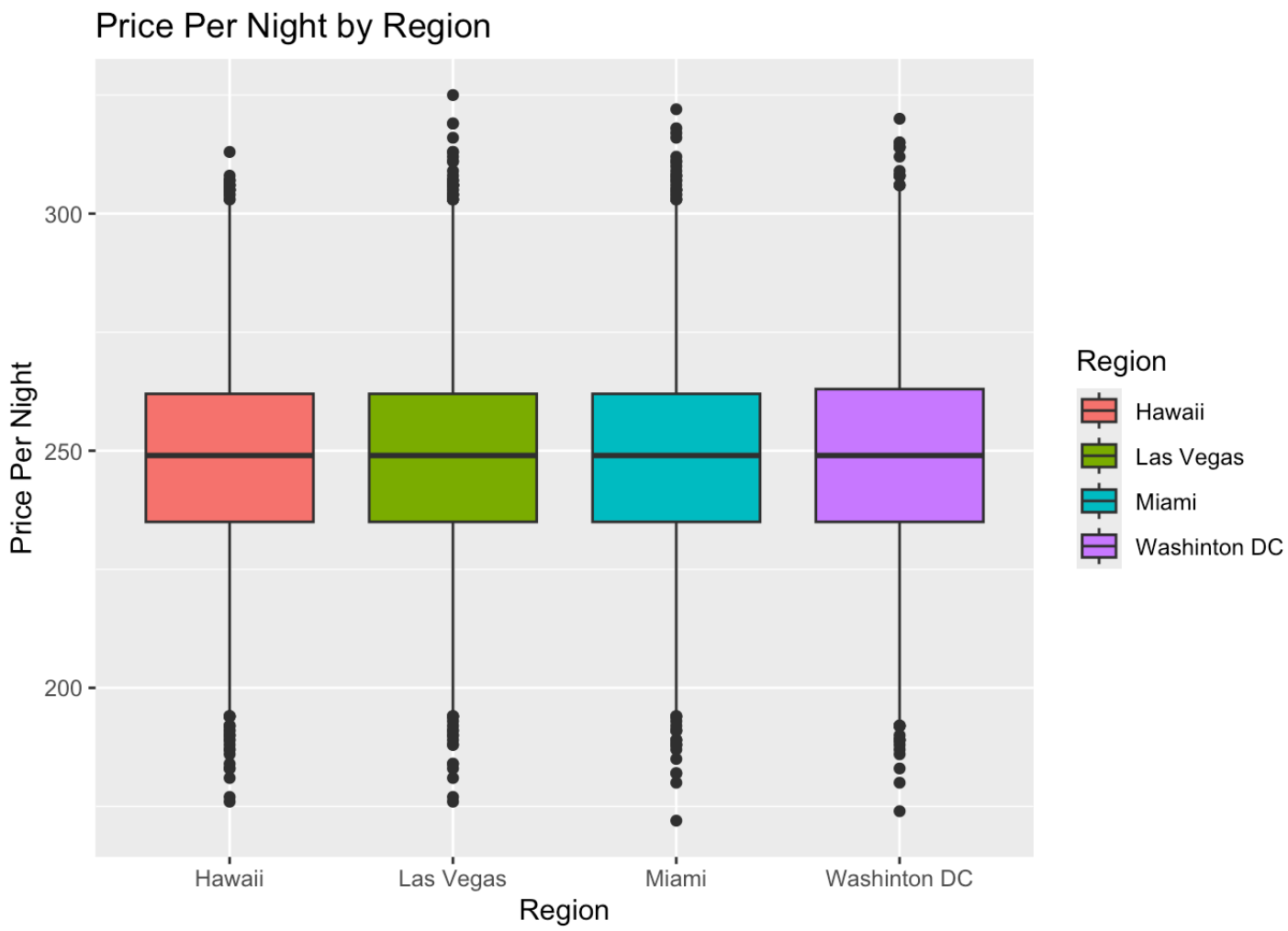

```
# Load the necessary library
library(scales)

# Histogram for User Income with formatted x-axis
ggplot(Expedia, aes(x = UserIncome)) +
  geom_histogram(bins = 30, fill = "green", color = "black") +
  scale_x_continuous(labels = comma) +
  labs(title = "Distribution of User Income", x = "User Income", y = "Frequency")
```



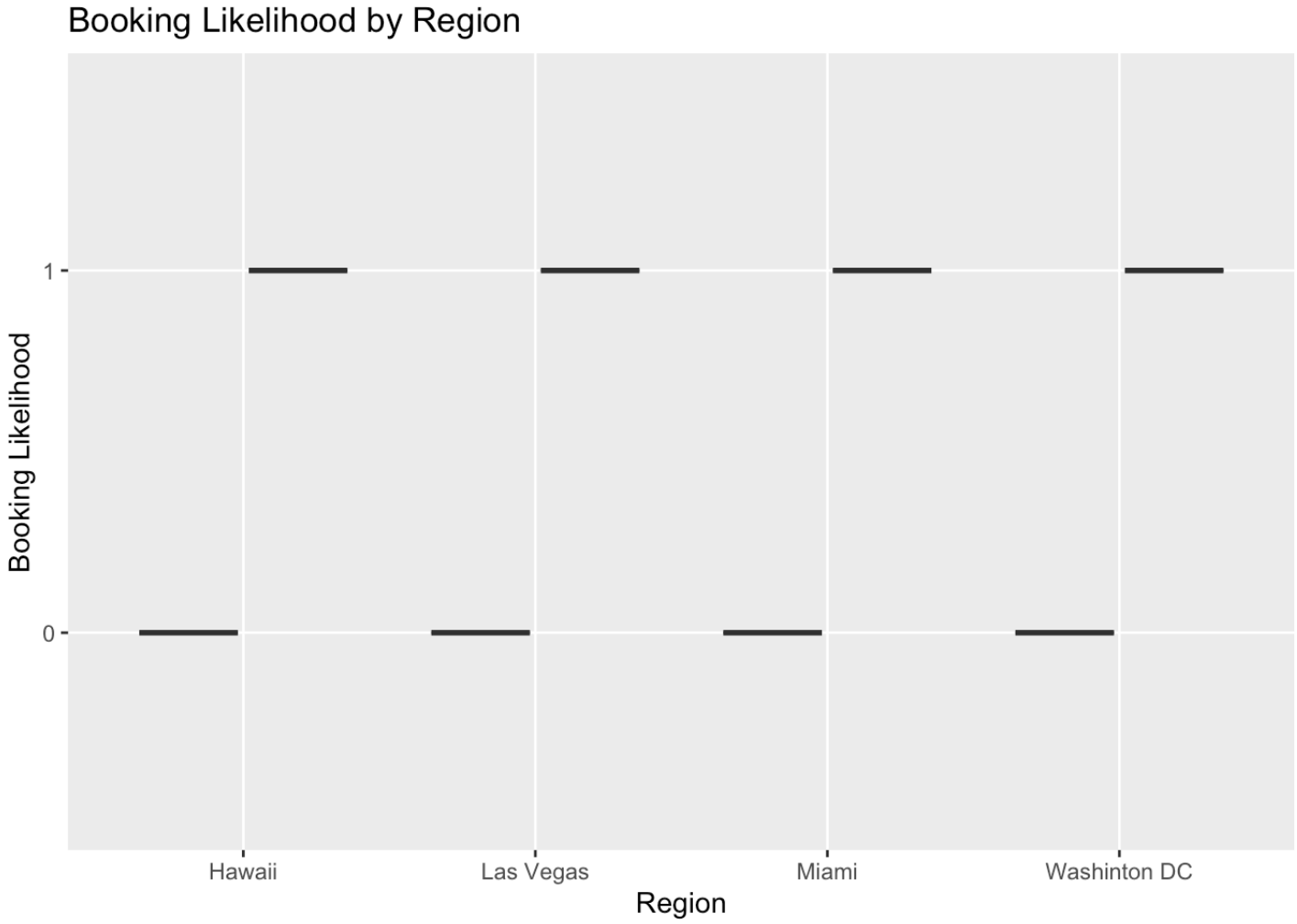

```
# Bar plot for Booking rates by Region
ggplot(Expedia, aes(x = Region, fill = as.factor(`Booked?`))) +
  geom_bar(position = "dodge") +
  labs(title = "Booking Rate by Region", x = "Region", y = "Count", fill = "Booked?")
```




```
# Boxplot for Price Per Night by Region
ggplot(Expedia, aes(x = Region, y = PricePerNight, fill = Region)) +
  geom_boxplot() +
  labs(title = "Price Per Night by Region", x = "Region", y = "Price Per Night")
```




```
# Boxplot for Booking Likelihood by Region
ggplot(Expedia, aes(x = Region, y = as.factor(`Booked?`))) +
  geom_boxplot() +
  labs(title = "Booking Likelihood by Region", x = "Region", y = "Booking Likelihoo
d")
```

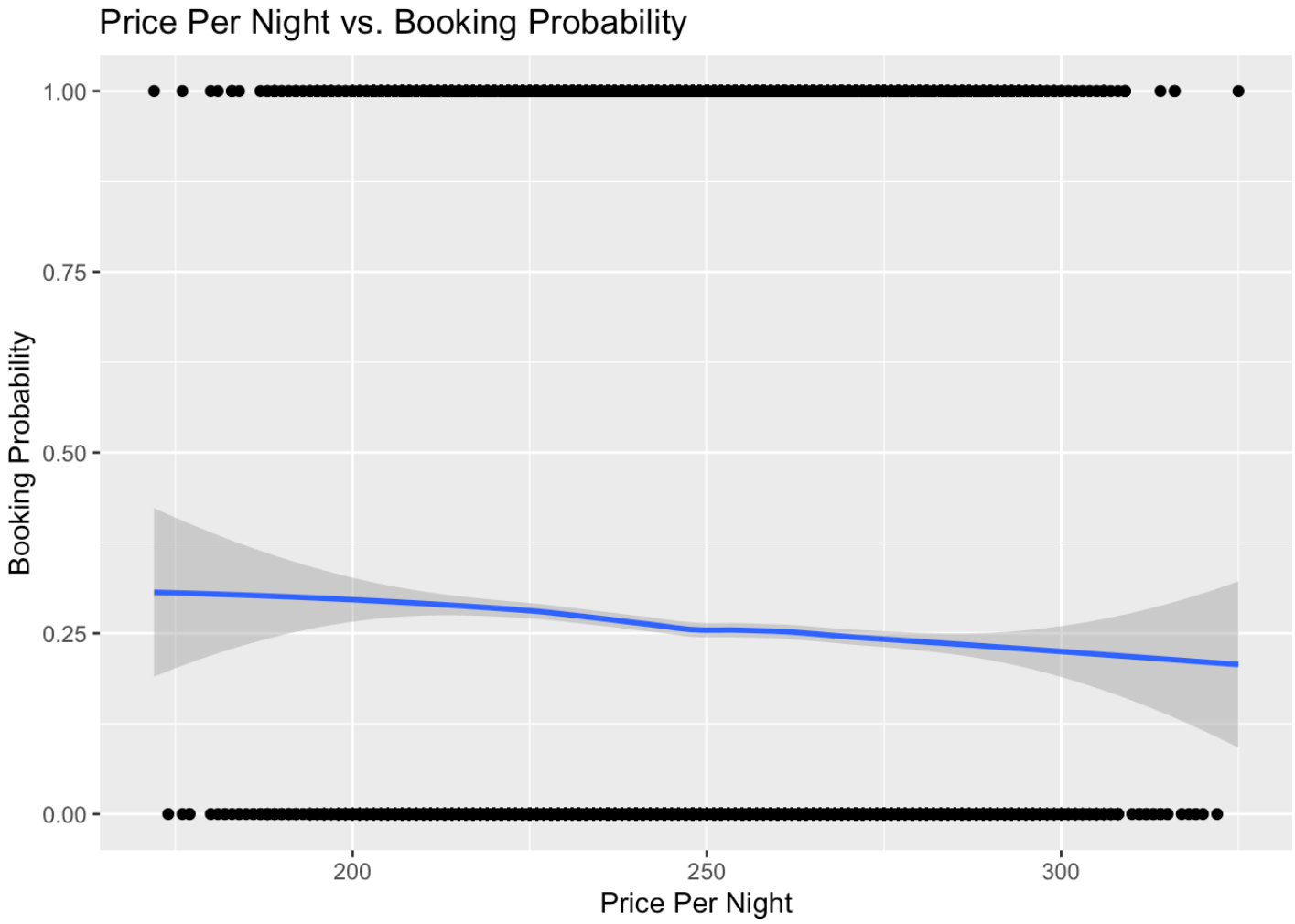



```
# Logistic Regression Model
model <- glm(`Booked?` ~ PricePerNight + UserIncome + Region +
             PricePerNight:UserIncome + PricePerNight:Region,
             data = Expedia, family = binomial)

# Summary of the model
summary(model)
```

```
##
## Call:
## glm(formula = `Booked?` ~ PricePerNight + UserIncome + Region +
##     PricePerNight:UserIncome + PricePerNight:Region, family = binomial,
##     data = Expedia)
##
## Coefficients:
##                                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)                    -1.280e-01  5.460e-01  -0.234 0.814700
## PricePerNight                  -9.802e-03  2.200e-03  -4.455 8.38e-06 ***
## UserIncome                     -2.997e-06  6.103e-06  -0.491 0.623369
## RegionLas Vegas                 9.920e-01  5.320e-01   1.864 0.062256 .
## RegionMiami                     1.113e+00  5.536e-01   2.010 0.044450 *
## RegionWashinton DC             -3.293e-01  5.538e-01  -0.595 0.552077
## PricePerNight:UserIncome        9.521e-08  2.450e-08   3.887 0.000102 ***
## PricePerNight:RegionLas Vegas  -6.413e-04  2.142e-03  -0.299 0.764584
## PricePerNight:RegionMiami      -3.319e-03  2.231e-03  -1.488 0.136832
## PricePerNight:RegionWashinton DC  2.492e-03  2.226e-03   1.119 0.263116
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 28640  on 24999  degrees of freedom
## Residual deviance: 26272  on 24990  degrees of freedom
## AIC: 26292
##
## Number of Fisher Scoring iterations: 4
```

```
# Scatter plot with smooth line for Booking Probability vs Price Per Night
ggplot(Expedia, aes(x = PricePerNight, y = as.numeric(`Booked?`))) +
  geom_point() +
  geom_smooth(method = "loess") +
  labs(title = "Price Per Night vs. Booking Probability", x = "Price Per Night", y =
"Booking Probability")
```

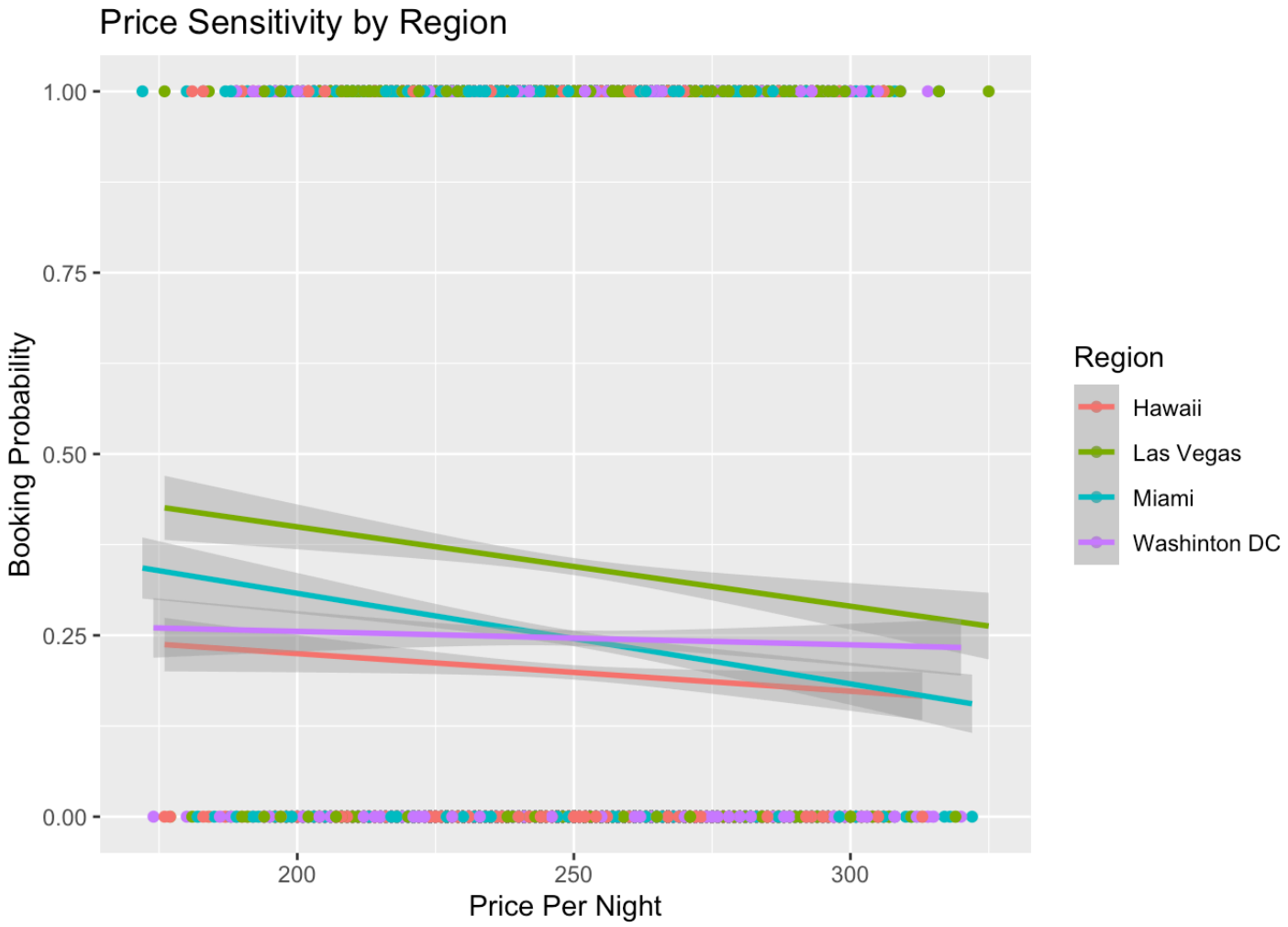
```
## `geom_smooth()` using formula = 'y ~ x'
```




```
### notes 2 slide 37
```

```
# Price Sensitivity by Region
ggplot(Expedia, aes(x = PricePerNight, y = as.numeric(`Booked?`), color = Region)) +
  geom_point() +
  geom_smooth(method = "lm") +
  labs(title = "Price Sensitivity by Region", x = "Price Per Night", y = "Booking Pro
bability")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

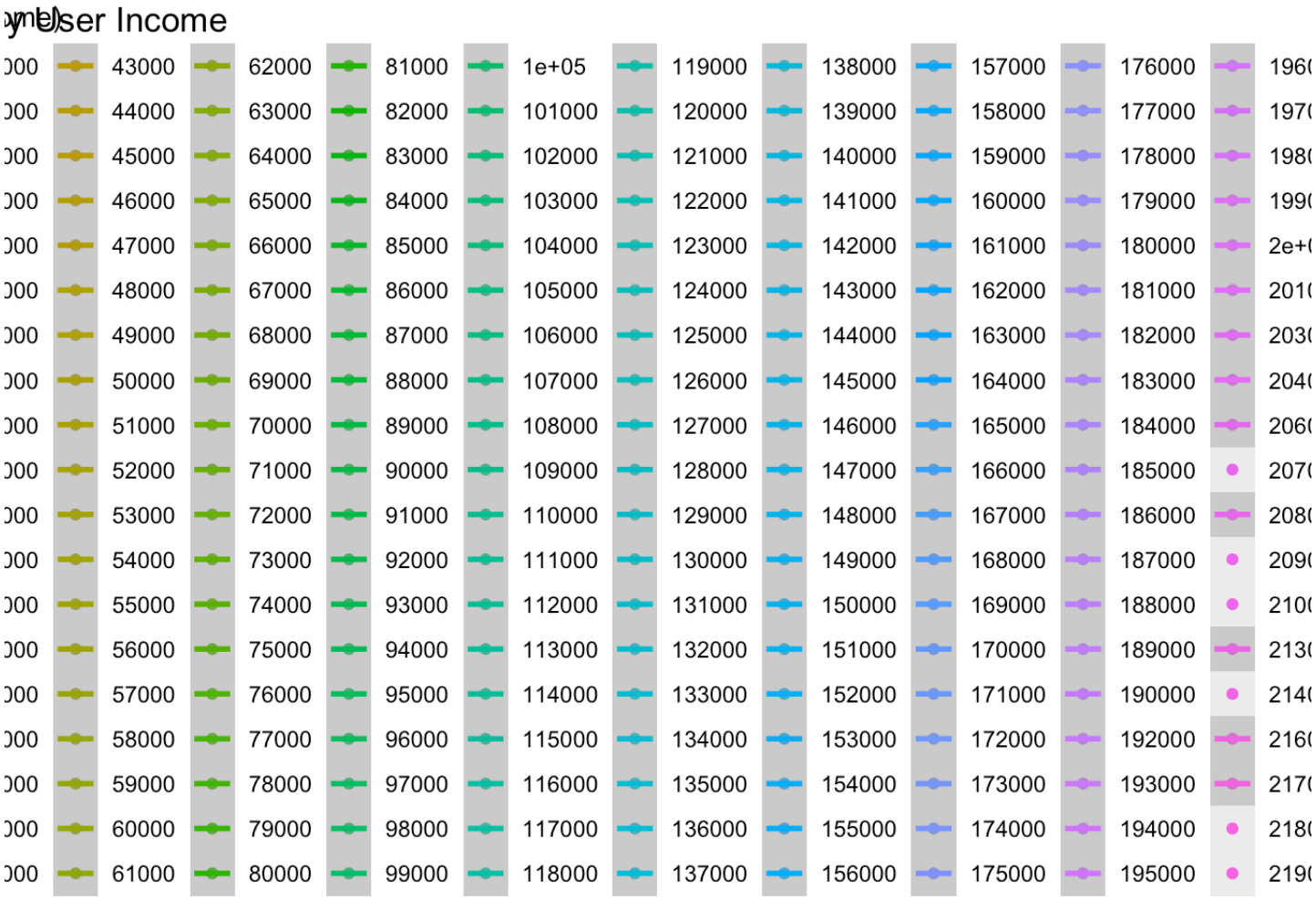



```
# Interaction between UserIncome and PricePerNight
ggplot(Expedia, aes(x = PricePerNight, y = as.numeric(`Booked?`), color = as.factor(U
serIncome))) +
  geom_point() +
  geom_smooth(method = "lm") +
  labs(title = "Price Sensitivity by User Income", x = "Price Per Night", y = "Bookin
g Probability")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

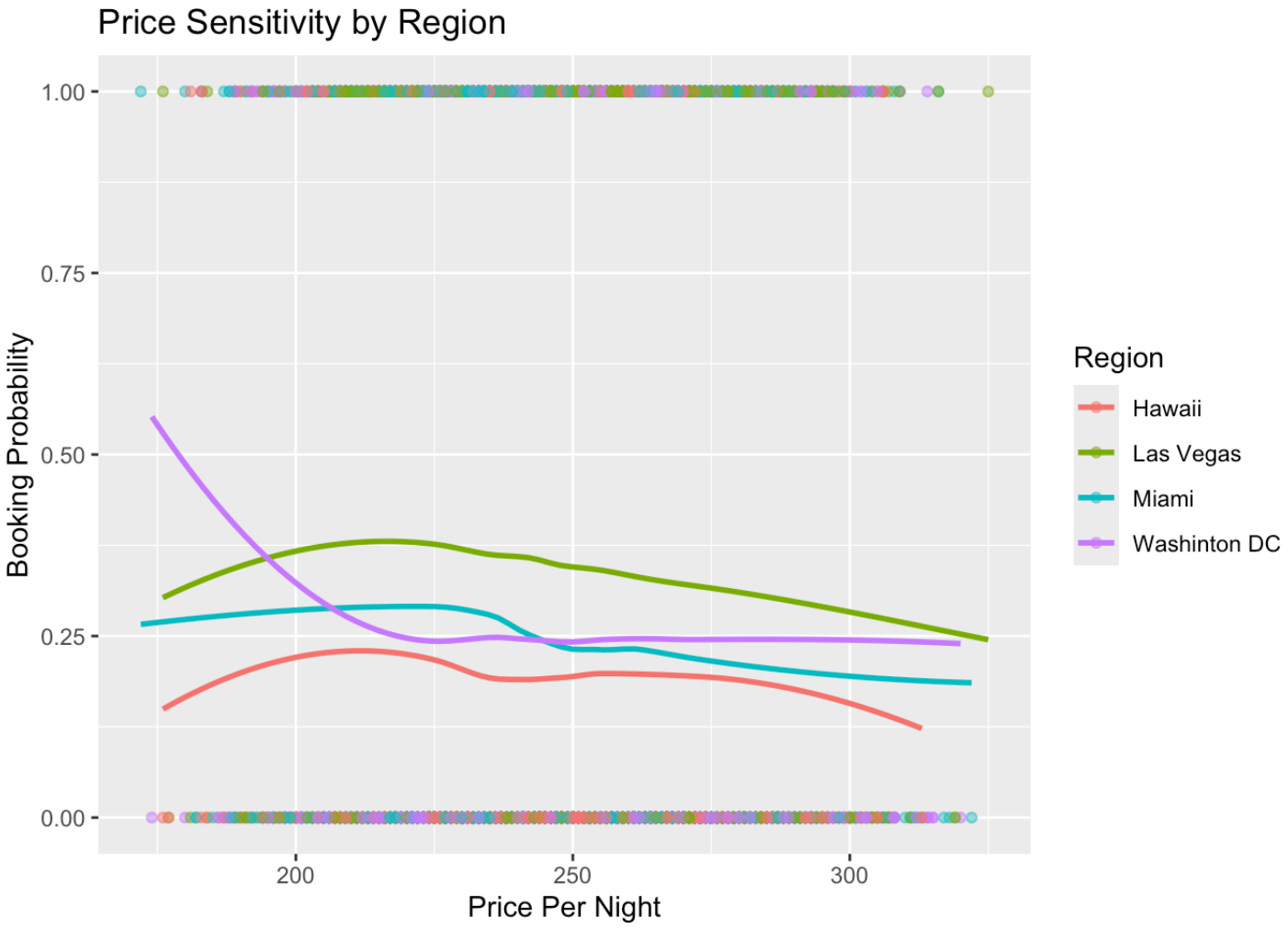
```
## Warning in qt((1 - level)/2, df): NaNs produced
## Warning in qt((1 - level)/2, df): NaNs produced
## Warning in qt((1 - level)/2, df): NaNs produced
## Warning in qt((1 - level)/2, df): NaNs produced
## Warning in qt((1 - level)/2, df): NaNs produced
## Warning in qt((1 - level)/2, df): NaNs produced
## Warning in qt((1 - level)/2, df): NaNs produced
## Warning in qt((1 - level)/2, df): NaNs produced
## Warning in qt((1 - level)/2, df): NaNs produced
## Warning in qt((1 - level)/2, df): NaNs produced
## Warning in qt((1 - level)/2, df): NaNs produced
## Warning in qt((1 - level)/2, df): NaNs produced
## Warning in qt((1 - level)/2, df): NaNs produced
## Warning in qt((1 - level)/2, df): NaNs produced
## Warning in qt((1 - level)/2, df): NaNs produced
## Warning in qt((1 - level)/2, df): NaNs produced
## Warning in qt((1 - level)/2, df): NaNs produced
```

```
## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf
## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf
## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf
## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf
## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf
## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf
## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf
## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf
## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf
## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf
## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf
## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf
## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf
## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf
## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf
## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf
## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf
```



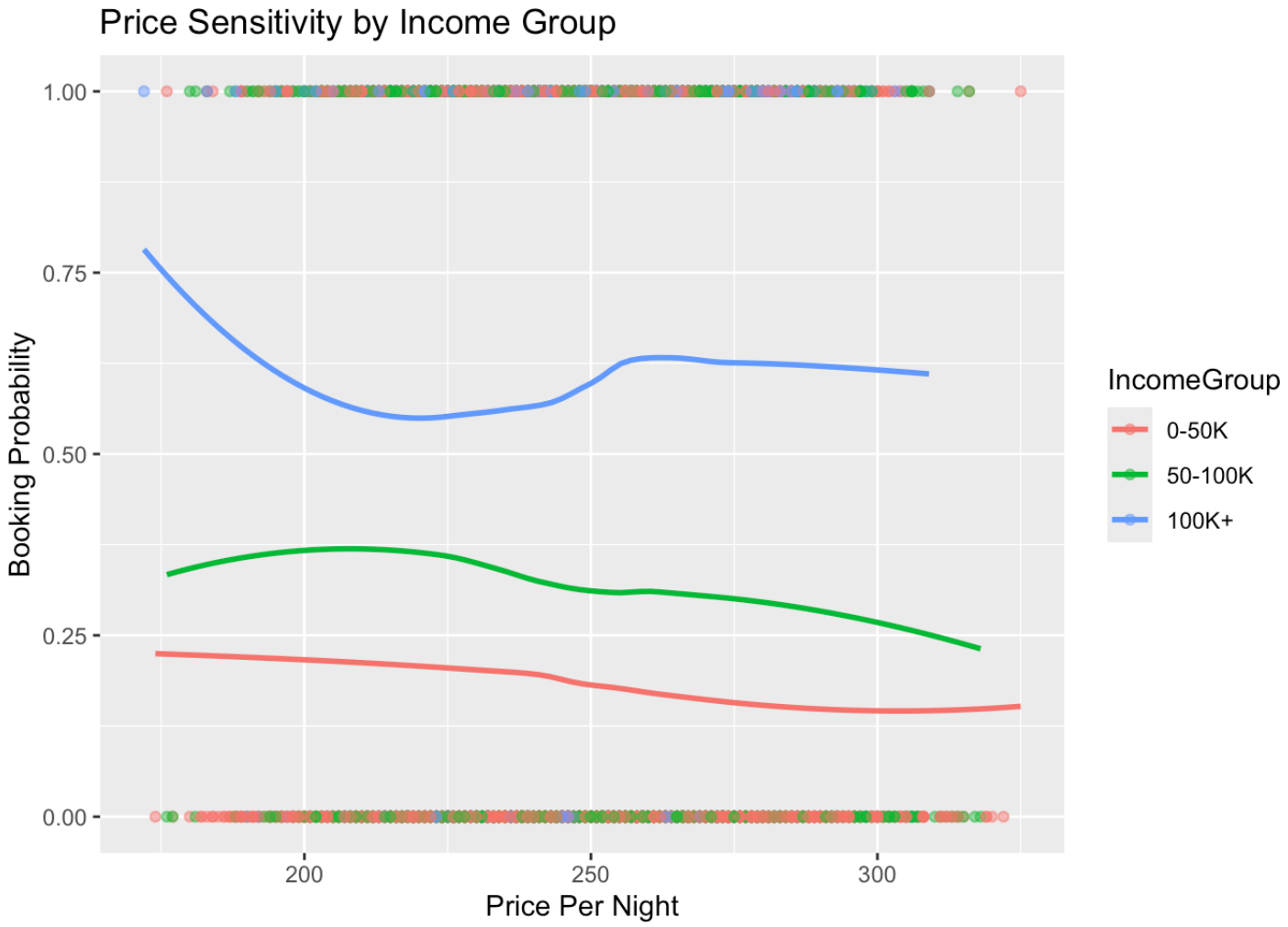

```
# Plotting Price Sensitivity by Region with adjusted y-axis and binning
ggplot(Expedia, aes(x = PricePerNight, y = as.numeric(`Booked?`), color = Region)) +
  geom_point(alpha = 0.5) +  # Adding transparency for better visualization
  geom_smooth(method = "loess", se = FALSE) +  # Use 'loess' to fit non-linear curves for better flexibility
  labs(title = "Price Sensitivity by Region", x = "Price Per Night", y = "Booking Probability")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
# Create bins for UserIncome
Expedia$IncomeGroup <- cut(Expedia$UserIncome,
                           breaks = c(0, 50000, 100000, Inf),
                           labels = c("0-50K", "50-100K", "100K+"))

# Plot using binned UserIncome groups
ggplot(Expedia, aes(x = PricePerNight, y = as.numeric(`Booked?`), color = IncomeGrou
p)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = "loess", se = FALSE) +
  labs(title = "Price Sensitivity by Income Group", x = "Price Per Night", y = "Booki
ng Probability")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
# Average booking rate by region
booking_rate <- Expedia %>%
  group_by(Region) %>%
  summarise(Avg_Booking = mean(`Booked?`))

print(booking_rate)
```

```
## # A tibble: 4 × 2
##   Region       Avg_Booking
##   <chr>              <dbl>
## 1 Hawaii             0.200
## 2 Las Vegas          0.346
## 3 Miami              0.247
## 4 Washinton DC       0.246
```

```
model1 <- lm(Expedia$`Booked?` ~ Expedia$PricePerNight * Expedia$UserIncome * Expedia
$Region, data = Expedia)
```

```
# Analysis 1: Price Elasticity on Booking Likelihood
# Linear regression for Booking Likelihood
model_booking <- lm(as.numeric(`Booked?`) ~ PricePerNight * UserIncome * Region, data
= Expedia)
summary(model_booking)
```

```
##
## Call:
## lm(formula = as.numeric(`Booked?`) ~ PricePerNight * UserIncome *
##     Region, data = Expedia)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.2990 -0.2661 -0.1717  0.3142  0.9755
##
## Coefficients:
##                                           Estimate Std. Error t value
## (Intercept)                              2.744e-01  1.287e-01   2.132
## PricePerNight                           -1.211e-03  5.157e-04  -2.348
## UserIncome                               1.419e-06  2.123e-06   0.668
## RegionLas Vegas                          2.034e-01  1.819e-01   1.118
## RegionMiami                              3.880e-01  1.809e-01   2.145
## RegionWashinton DC                      -3.563e-02  1.797e-01  -0.198
## PricePerNight:UserIncome                 1.178e-08  8.494e-09   1.387
## PricePerNight:RegionLas Vegas           -1.425e-04  7.287e-04  -0.196
## PricePerNight:RegionMiami               -1.276e-03  7.243e-04  -1.761
## PricePerNight:RegionWashinton DC         4.127e-04  7.189e-04   0.574
## UserIncome:RegionLas Vegas               1.397e-06  3.029e-06   0.461
## UserIncome:RegionMiami                  -3.270e-06  2.956e-06  -1.106
## UserIncome:RegionWashinton DC            1.404e-07  2.944e-06   0.048
## PricePerNight:UserIncome:RegionLas Vegas -7.091e-09  1.212e-08  -0.585
## PricePerNight:UserIncome:RegionMiami     1.128e-08  1.182e-08   0.954
## PricePerNight:UserIncome:RegionWashinton DC -2.196e-09  1.176e-08  -0.187
##                                            Pr(>|t|)
## (Intercept)                                  0.0330 *
## PricePerNight                                0.0189 *
## UserIncome                                   0.5041
## RegionLas Vegas                              0.2636
## RegionMiami                                  0.0320 *
## RegionWashinton DC                           0.8428
## PricePerNight:UserIncome                     0.1654
## PricePerNight:RegionLas Vegas                0.8449
## PricePerNight:RegionMiami                    0.0782 .
## PricePerNight:RegionWashinton DC             0.5660
## UserIncome:RegionLas Vegas                   0.6446
## UserIncome:RegionMiami                       0.2685
```

```
## UserIncome:RegionWashinton DC                 0.9620
## PricePerNight:UserIncome:RegionLas Vegas      0.5586
## PricePerNight:UserIncome:RegionMiami          0.3400
## PricePerNight:UserIncome:RegionWashinton DC   0.8518
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4165 on 24984 degrees of freedom
## Multiple R-squared:  0.09847,    Adjusted R-squared:  0.09793
## F-statistic: 181.9 on 15 and 24984 DF,  p-value: < 2.2e-16
```

```
# Extracting key results
coef_summary <- summary(model_booking)$coefficients
coef_summary
```

```
##                                                 Estimate   Std. Error
## (Intercept)                                   2.744125e-01 1.287264e-01
## PricePerNight                                -1.211007e-03 5.156963e-04
## UserIncome                                    1.418553e-06 2.123345e-06
## RegionLas Vegas                               2.033620e-01 1.819046e-01
## RegionMiami                                   3.879897e-01 1.809108e-01
## RegionWashinton DC                           -3.563337e-02 1.797192e-01
## PricePerNight:UserIncome                      1.178189e-08 8.494073e-09
## PricePerNight:RegionLas Vegas                -1.425287e-04 7.286869e-04
## PricePerNight:RegionMiami                    -1.275722e-03 7.242803e-04
## PricePerNight:RegionWashinton DC              4.126600e-04 7.189388e-04
## UserIncome:RegionLas Vegas                    1.397326e-06 3.028855e-06
## UserIncome:RegionMiami                       -3.270474e-06 2.955807e-06
## UserIncome:RegionWashinton DC                 1.404265e-07 2.944421e-06
## PricePerNight:UserIncome:RegionLas Vegas     -7.090673e-09 1.212345e-08
## PricePerNight:UserIncome:RegionMiami          1.128057e-08 1.182137e-08
## PricePerNight:UserIncome:RegionWashinton DC -2.196442e-09 1.175642e-08
##                                                  t value   Pr(>|t|)
## (Intercept)                                   2.13175056 0.03303709
## PricePerNight                                -2.34829393 0.01886738
## UserIncome                                    0.66807445 0.50409224
## RegionLas Vegas                               1.11795919 0.26359517
## RegionMiami                                   2.14464684 0.03199071
## RegionWashinton DC                           -0.19827247 0.84283350
## PricePerNight:UserIncome                      1.38707179 0.16543225
## PricePerNight:RegionLas Vegas                -0.19559655 0.84492755
## PricePerNight:RegionMiami                    -1.76136462 0.07818894
## PricePerNight:RegionWashinton DC              0.57398485 0.56598322
## UserIncome:RegionLas Vegas                    0.46133804 0.64456012
## UserIncome:RegionMiami                       -1.10645730 0.26853929
## UserIncome:RegionWashinton DC                 0.04769239 0.96196178
## PricePerNight:UserIncome:RegionLas Vegas     -0.58487242 0.55863878
## PricePerNight:UserIncome:RegionMiami          0.95425266 0.33996500
## PricePerNight:UserIncome:RegionWashinton DC -0.18682917 0.85179609
```

```
#Analysis 2: Price Elasticity on Number of Nights Booked

# Filter the data for booked cases only
booked_data <- Expedia %>% filter(`Booked?` == 1)

# Linear regression for Number of Nights Booked
model_nights <- lm(Nights ~ PricePerNight * UserIncome * Region, data = booked_data)
summary(model_nights)
```

```
##
```

```
## Call:
## lm(formula = Nights ~ PricePerNight * UserIncome * Region, data = booked_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.15734 -0.84598  0.03951  0.14483  3.03037
##
## Coefficients:
##                                          Estimate Std. Error t value
## (Intercept)                             4.178e+00  5.427e-01   7.698
## PricePerNight                          -4.913e-03  2.193e-03  -2.240
## UserIncome                             -5.605e-06  6.531e-06  -0.858
## RegionLas Vegas                        -1.327e+00  6.747e-01  -1.967
## RegionMiami                            -5.712e-01  7.165e-01  -0.797
## RegionWashinton DC                     -7.109e-01  7.006e-01  -1.015
## PricePerNight:UserIncome                2.162e-08  2.624e-08   0.824
## PricePerNight:RegionLas Vegas           5.497e-03  2.723e-03   2.019
## PricePerNight:RegionMiami               2.346e-03  2.896e-03   0.810
## PricePerNight:RegionWashinton DC        2.698e-03  2.819e-03   0.957
## UserIncome:RegionLas Vegas              1.790e-05  8.605e-06   2.080
## UserIncome:RegionMiami                  6.981e-06  8.679e-06   0.804
## UserIncome:RegionWashinton DC           3.266e-06  8.630e-06   0.378
## PricePerNight:UserIncome:RegionLas Vegas -7.222e-08  3.456e-08  -2.090
## PricePerNight:UserIncome:RegionMiami    -2.870e-08  3.484e-08  -0.824
## PricePerNight:UserIncome:RegionWashinton DC -1.130e-08  3.451e-08  -0.327
##                                         Pr(>|t|)
## (Intercept)                             1.59e-14 ***
## PricePerNight                             0.0251 *
## UserIncome                                0.3908
## RegionLas Vegas                           0.0492 *
## RegionMiami                               0.4253
## RegionWashinton DC                        0.3103
## PricePerNight:UserIncome                  0.4100
## PricePerNight:RegionLas Vegas             0.0435 *
## PricePerNight:RegionMiami                 0.4179
## PricePerNight:RegionWashinton DC          0.3385
## UserIncome:RegionLas Vegas                0.0376 *
## UserIncome:RegionMiami                    0.4212
## UserIncome:RegionWashinton DC             0.7051
## PricePerNight:UserIncome:RegionLas Vegas  0.0367 *
## PricePerNight:UserIncome:RegionMiami      0.4101
## PricePerNight:UserIncome:RegionWashinton DC  0.7434
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7547 on 6478 degrees of freedom
## Multiple R-squared:  0.006935,   Adjusted R-squared:  0.004635
```

```
## F-statistic: 3.016 on 15 and 6478 DF,  p-value: 7.314e-05
```

```
# Extracting key results
coef_summary_nights <- summary(model_nights)$coefficients
coef_summary_nights
```

```
##                                              Estimate   Std. Error
## (Intercept)                                4.177671e+00 5.426934e-01
## PricePerNight                             -4.913229e-03 2.193271e-03
## UserIncome                                -5.605284e-06 6.531054e-06
## RegionLas Vegas                           -1.327304e+00 6.747345e-01
## RegionMiami                               -5.712488e-01 7.164906e-01
## RegionWashinton DC                        -7.108718e-01 7.005597e-01
## PricePerNight:UserIncome                   2.162330e-08 2.624402e-08
## PricePerNight:RegionLas Vegas              5.497390e-03 2.722745e-03
## PricePerNight:RegionMiami                  2.346458e-03 2.896169e-03
## PricePerNight:RegionWashinton DC           2.698265e-03 2.819003e-03
## UserIncome:RegionLas Vegas                 1.789612e-05 8.605204e-06
## UserIncome:RegionMiami                     6.981088e-06 8.678780e-06
## UserIncome:RegionWashinton DC              3.266361e-06 8.630273e-06
## PricePerNight:UserIncome:RegionLas Vegas  -7.222158e-08 3.455835e-08
## PricePerNight:UserIncome:RegionMiami      -2.869669e-08 3.483712e-08
## PricePerNight:UserIncome:RegionWashinton DC -1.129711e-08 3.450699e-08
##                                               t value     Pr(>|t|)
## (Intercept)                                 7.6980325 1.588182e-14
## PricePerNight                              -2.2401373 2.511579e-02
## UserIncome                                 -0.8582510 3.907856e-01
## RegionLas Vegas                            -1.9671499 4.920854e-02
## RegionMiami                                -0.7972872 4.253135e-01
## RegionWashinton DC                         -1.0147197 3.102774e-01
## PricePerNight:UserIncome                    0.8239326 4.100082e-01
## PricePerNight:RegionLas Vegas               2.0190614 4.352195e-02
## PricePerNight:RegionMiami                   0.8101937 4.178586e-01
## PricePerNight:RegionWashinton DC            0.9571699 3.385172e-01
## UserIncome:RegionLas Vegas                  2.0796851 3.759365e-02
## UserIncome:RegionMiami                      0.8043859 4.212036e-01
## UserIncome:RegionWashinton DC               0.3784771 7.050886e-01
## PricePerNight:UserIncome:RegionLas Vegas   -2.0898446 3.667066e-02
## PricePerNight:UserIncome:RegionMiami       -0.8237388 4.101183e-01
## PricePerNight:UserIncome:RegionWashinton DC -0.3273862 7.433864e-01
```

```
# Summary of key findings
summary_booking <- summary(model_booking)
summary_nights <- summary(model_nights)

# Compare coefficients and key statistics
summary_booking$coefficients
```

```
##                                              Estimate   Std. Error
## (Intercept)                                 2.744125e-01 1.287264e-01
## PricePerNight                              -1.211007e-03 5.156963e-04
## UserIncome                                  1.418553e-06 2.123345e-06
## RegionLas Vegas                             2.033620e-01 1.819046e-01
## RegionMiami                                 3.879897e-01 1.809108e-01
## RegionWashinton DC                         -3.563337e-02 1.797192e-01
## PricePerNight:UserIncome                    1.178189e-08 8.494073e-09
## PricePerNight:RegionLas Vegas              -1.425287e-04 7.286869e-04
## PricePerNight:RegionMiami                  -1.275722e-03 7.242803e-04
## PricePerNight:RegionWashinton DC            4.126600e-04 7.189388e-04
## UserIncome:RegionLas Vegas                  1.397326e-06 3.028855e-06
## UserIncome:RegionMiami                     -3.270474e-06 2.955807e-06
## UserIncome:RegionWashinton DC               1.404265e-07 2.944421e-06
## PricePerNight:UserIncome:RegionLas Vegas   -7.090673e-09 1.212345e-08
## PricePerNight:UserIncome:RegionMiami        1.128057e-08 1.182137e-08
## PricePerNight:UserIncome:RegionWashinton DC -2.196442e-09 1.175642e-08
##                                                t value   Pr(>|t|)
## (Intercept)                                 2.13175056 0.03303709
## PricePerNight                              -2.34829393 0.01886738
## UserIncome                                  0.66807445 0.50409224
## RegionLas Vegas                             1.11795919 0.26359517
## RegionMiami                                 2.14464684 0.03199071
## RegionWashinton DC                         -0.19827247 0.84283350
## PricePerNight:UserIncome                    1.38707179 0.16543225
## PricePerNight:RegionLas Vegas              -0.19559655 0.84492755
## PricePerNight:RegionMiami                  -1.76136462 0.07818894
## PricePerNight:RegionWashinton DC            0.57398485 0.56598322
## UserIncome:RegionLas Vegas                  0.46133804 0.64456012
## UserIncome:RegionMiami                     -1.10645730 0.26853929
## UserIncome:RegionWashinton DC               0.04769239 0.96196178
## PricePerNight:UserIncome:RegionLas Vegas   -0.58487242 0.55863878
## PricePerNight:UserIncome:RegionMiami        0.95425266 0.33996500
## PricePerNight:UserIncome:RegionWashinton DC -0.18682917 0.85179609
```

```
summary_nights$coefficients
```

```
##                                                Estimate   Std. Error
## (Intercept)                                4.177671e+00 5.426934e-01
## PricePerNight                             -4.913229e-03 2.193271e-03
## UserIncome                                -5.605284e-06 6.531054e-06
## RegionLas Vegas                           -1.327304e+00 6.747345e-01
## RegionMiami                               -5.712488e-01 7.164906e-01
## RegionWashinton DC                        -7.108718e-01 7.005597e-01
## PricePerNight:UserIncome                   2.162330e-08 2.624402e-08
## PricePerNight:RegionLas Vegas              5.497390e-03 2.722745e-03
## PricePerNight:RegionMiami                  2.346458e-03 2.896169e-03
## PricePerNight:RegionWashinton DC           2.698265e-03 2.819003e-03
## UserIncome:RegionLas Vegas                 1.789612e-05 8.605204e-06
## UserIncome:RegionMiami                     6.981088e-06 8.678780e-06
## UserIncome:RegionWashinton DC              3.266361e-06 8.630273e-06
## PricePerNight:UserIncome:RegionLas Vegas  -7.222158e-08 3.455835e-08
## PricePerNight:UserIncome:RegionMiami      -2.869669e-08 3.483712e-08
## PricePerNight:UserIncome:RegionWashinton DC -1.129711e-08 3.450699e-08
##                                              t value     Pr(>|t|)
## (Intercept)                                7.6980325 1.588182e-14
## PricePerNight                             -2.2401373 2.511579e-02
## UserIncome                                -0.8582510 3.907856e-01
## RegionLas Vegas                           -1.9671499 4.920854e-02
## RegionMiami                               -0.7972872 4.253135e-01
## RegionWashinton DC                        -1.0147197 3.102774e-01
## PricePerNight:UserIncome                   0.8239326 4.100082e-01
## PricePerNight:RegionLas Vegas              2.0190614 4.352195e-02
## PricePerNight:RegionMiami                  0.8101937 4.178586e-01
## PricePerNight:RegionWashinton DC           0.9571699 3.385172e-01
## UserIncome:RegionLas Vegas                 2.0796851 3.759365e-02
## UserIncome:RegionMiami                     0.8043859 4.212036e-01
## UserIncome:RegionWashinton DC              0.3784771 7.050886e-01
## PricePerNight:UserIncome:RegionLas Vegas  -2.0898446 3.667066e-02
## PricePerNight:UserIncome:RegionMiami      -0.8237388 4.101183e-01
## PricePerNight:UserIncome:RegionWashinton DC -0.3273862 7.433864e-01
```

```
# Logistic regression for Booking Likelihood
model_booking_logistic <- glm(as.numeric(`Booked?`) ~ PricePerNight * UserIncome * Re
gion,
                              data = Expedia,
                              family = binomial)

# Summary of the logistic model
summary(model_booking_logistic)
```

```
##
```

```
## Call:
## glm(formula = as.numeric(`Booked?`) ~ PricePerNight * UserIncome *
##     Region, family = binomial, data = Expedia)
##
## Coefficients:
##                                          Estimate Std. Error z value
## (Intercept)                             -2.569e-01  8.947e-01  -0.287
## PricePerNight                           -1.029e-02  3.608e-03  -2.851
## UserIncome                               3.886e-07  1.309e-05   0.030
## RegionLas Vegas                          4.492e-01  1.147e+00   0.392
## RegionMiami                              2.362e+00  1.196e+00   1.975
## RegionWashinton DC                      -4.567e-01  1.198e+00  -0.381
## PricePerNight:UserIncome                 9.826e-08  5.260e-08   1.868
## PricePerNight:RegionLas Vegas            3.095e-03  4.620e-03   0.670
## PricePerNight:RegionMiami               -7.244e-03  4.829e-03  -1.500
## PricePerNight:RegionWashinton DC         4.137e-03  4.822e-03   0.858
## UserIncome:RegionLas Vegas               8.783e-06  1.769e-05   0.497
## UserIncome:RegionMiami                  -2.261e-05  1.769e-05  -1.278
## UserIncome:RegionWashinton DC            8.995e-07  1.783e-05   0.050
## PricePerNight:UserIncome:RegionLas Vegas -6.221e-08  7.100e-08  -0.876
## PricePerNight:UserIncome:RegionMiami      7.246e-08  7.116e-08   1.018
## PricePerNight:UserIncome:RegionWashinton DC -2.246e-08  7.152e-08  -0.314
##                                          Pr(>|z|)
## (Intercept)                               0.77401
## PricePerNight                             0.00435 **
## UserIncome                                0.97632
## RegionLas Vegas                           0.69537
## RegionMiami                               0.04832 *
## RegionWashinton DC                        0.70309
## PricePerNight:UserIncome                  0.06176 .
## PricePerNight:RegionLas Vegas             0.50286
## PricePerNight:RegionMiami                 0.13362
## PricePerNight:RegionWashinton DC          0.39102
## UserIncome:RegionLas Vegas                0.61950
## UserIncome:RegionMiami                    0.20136
## UserIncome:RegionWashinton DC             0.95976
## PricePerNight:UserIncome:RegionLas Vegas  0.38092
## PricePerNight:UserIncome:RegionMiami      0.30853
## PricePerNight:UserIncome:RegionWashinton DC  0.75346
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 28640  on 24999  degrees of freedom
## Residual deviance: 26246  on 24984  degrees of freedom
## AIC: 26278
```

```
##
## Number of Fisher Scoring iterations: 4
```

```
# Extracting key results
coef_summary_logistic <- summary(model_booking_logistic)$coefficients
coef_summary_logistic
```

```
##                                              Estimate   Std. Error
## (Intercept)                                -2.568978e-01 8.947067e-01
## PricePerNight                              -1.028848e-02 3.608306e-03
## UserIncome                                  3.886500e-07 1.309360e-05
## RegionLas Vegas                             4.491749e-01 1.147072e+00
## RegionMiami                                 2.361769e+00 1.196113e+00
## RegionWashinton DC                         -4.567046e-01 1.198211e+00
## PricePerNight:UserIncome                    9.826126e-08 5.260248e-08
## PricePerNight:RegionLas Vegas               3.095106e-03 4.619603e-03
## PricePerNight:RegionMiami                  -7.244116e-03 4.829428e-03
## PricePerNight:RegionWashinton DC            4.136555e-03 4.822487e-03
## UserIncome:RegionLas Vegas                  8.783232e-06 1.768795e-05
## UserIncome:RegionMiami                     -2.260765e-05 1.769437e-05
## UserIncome:RegionWashinton DC               8.994935e-07 1.782887e-05
## PricePerNight:UserIncome:RegionLas Vegas   -6.221014e-08 7.099936e-08
## PricePerNight:UserIncome:RegionMiami        7.246059e-08 7.115703e-08
## PricePerNight:UserIncome:RegionWashinton DC -2.246131e-08 7.151529e-08
##                                                z value    Pr(>|z|)
## (Intercept)                                -0.28713076 0.774012199
## PricePerNight                              -2.85133368 0.004353626
## UserIncome                                  0.02968243 0.976320321
## RegionLas Vegas                             0.39158387 0.695365706
## RegionMiami                                 1.97453573 0.048320859
## RegionWashinton DC                         -0.38115549 0.703087873
## PricePerNight:UserIncome                    1.86799684 0.061762512
## PricePerNight:RegionLas Vegas               0.66999387 0.502861695
## PricePerNight:RegionMiami                  -1.49999482 0.133615745
## PricePerNight:RegionWashinton DC            0.85776378 0.391022914
## UserIncome:RegionLas Vegas                  0.49656573 0.619495328
## UserIncome:RegionMiami                     -1.27767450 0.201364218
## UserIncome:RegionWashinton DC               0.05045153 0.959762575
## PricePerNight:UserIncome:RegionLas Vegas   -0.87620707 0.380917474
## PricePerNight:UserIncome:RegionMiami        1.01831948 0.308526152
## PricePerNight:UserIncome:RegionWashinton DC -0.31407709 0.753462487
```

```
exp(coef(model_booking_logistic))
```

```
##                                   (Intercept)
##                                     0.7734473
##                                PricePerNight
##                                     0.9897643
##                                    UserIncome
##                                     1.0000004
##                               RegionLas Vegas
##                                     1.5670188
##                                  RegionMiami
##                                    10.6096986
##                             RegionWashinton DC
##                                     0.6333674
##                     PricePerNight:UserIncome
##                                     1.0000001
##                PricePerNight:RegionLas Vegas
##                                     1.0030999
##                    PricePerNight:RegionMiami
##                                     0.9927821
##             PricePerNight:RegionWashinton DC
##                                     1.0041451
##                   UserIncome:RegionLas Vegas
##                                     1.0000088
##                       UserIncome:RegionMiami
##                                     0.9999774
##                UserIncome:RegionWashinton DC
##                                     1.0000009
##    PricePerNight:UserIncome:RegionLas Vegas
##                                     0.9999999
##        PricePerNight:UserIncome:RegionMiami
##                                     1.0000001
## PricePerNight:UserIncome:RegionWashinton DC
##                                     1.0000000
```

```r
# Load necessary libraries
library(dplyr)

# Load the data
load("casedata.Rdata")

# Check the structure of the data
str(Expedia)
```

```
## 'data.frame':    25000 obs. of  5 variables:
##  $ PricePerNight: num  253 241 272 241 264 259 260 261 250 265 ...
##  $ Region       : chr  "Las Vegas" "Hawaii" "Miami" "Washinton DC" ...
##  $ UserIncome   : num  32000 49000 28000 107000 54000 28000 58000 67000 61000 3800
0 ...
##  $ Booked?      : num  0 0 0 1 1 0 1 1 0 1 ...
##  $ Nights       : num  0 0 0 2 3 0 4 3 0 4 ...
```

```
# Simple Linear Regression for Booking Likelihood
model_booking_simple <- lm(as.numeric(`Booked?`) ~ PricePerNight, data = Expedia)
summary(model_booking_simple)
```

```
##
## Call:
## lm(formula = as.numeric(`Booked?`) ~ PricePerNight, data = Expedia)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.3159 -0.2657 -0.2522  0.7126  0.7973
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.4463765  0.0342230  13.043  < 2e-16 ***
## PricePerNight -0.0007498  0.0001371  -5.471 4.52e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4383 on 24998 degrees of freedom
## Multiple R-squared:  0.001196,   Adjusted R-squared:  0.001156
## F-statistic: 29.93 on 1 and 24998 DF,  p-value: 4.52e-08
```

```
# Filter the data for booked cases only
booked_data <- Expedia %>% filter(`Booked?` == 1)

# Simple Linear Regression for Number of Nights Booked
model_nights_simple <- lm(Nights ~ PricePerNight, data = booked_data)
summary(model_nights_simple)
```

```
##
## Call:
## lm(formula = Nights ~ PricePerNight, data = booked_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.10881 -0.86191  0.04193  0.13029  3.04972
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    3.5974129  0.1149833  31.286  < 2e-16 ***
## PricePerNight -0.0025989  0.0004627  -5.617 2.02e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7546 on 6492 degrees of freedom
## Multiple R-squared:  0.004837,   Adjusted R-squared:  0.004684
## F-statistic: 31.56 on 1 and 6492 DF,  p-value: 2.019e-08
```

```
# Linear Regression for Booking Likelihood with interactions
model_booking_het <- lm(as.numeric(`Booked?`) ~ PricePerNight * UserIncome * Region,
data = Expedia)
summary(model_booking_het)
```

```
##
## Call:
## lm(formula = as.numeric(`Booked?`) ~ PricePerNight * UserIncome *
##     Region, data = Expedia)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.2990 -0.2661 -0.1717  0.3142  0.9755
##
## Coefficients:
##                                          Estimate Std. Error t value
## (Intercept)                             2.744e-01  1.287e-01   2.132
## PricePerNight                          -1.211e-03  5.157e-04  -2.348
## UserIncome                              1.419e-06  2.123e-06   0.668
## RegionLas Vegas                         2.034e-01  1.819e-01   1.118
## RegionMiami                             3.880e-01  1.809e-01   2.145
## RegionWashinton DC                     -3.563e-02  1.797e-01  -0.198
## PricePerNight:UserIncome                1.178e-08  8.494e-09   1.387
## PricePerNight:RegionLas Vegas          -1.425e-04  7.287e-04  -0.196
## PricePerNight:RegionMiami              -1.276e-03  7.243e-04  -1.761
## PricePerNight:RegionWashinton DC        4.127e-04  7.189e-04   0.574
```

```
## UserIncome:RegionLas Vegas                     1.397e-06  3.029e-06   0.461
## UserIncome:RegionMiami                        -3.270e-06  2.956e-06  -1.106
## UserIncome:RegionWashinton DC                  1.404e-07  2.944e-06   0.048
## PricePerNight:UserIncome:RegionLas Vegas      -7.091e-09  1.212e-08  -0.585
## PricePerNight:UserIncome:RegionMiami           1.128e-08  1.182e-08   0.954
## PricePerNight:UserIncome:RegionWashinton DC -2.196e-09  1.176e-08  -0.187
##                                               Pr(>|t|)
## (Intercept)                                     0.0330 *
## PricePerNight                                   0.0189 *
## UserIncome                                      0.5041
## RegionLas Vegas                                 0.2636
## RegionMiami                                     0.0320 *
## RegionWashinton DC                              0.8428
## PricePerNight:UserIncome                        0.1654
## PricePerNight:RegionLas Vegas                   0.8449
## PricePerNight:RegionMiami                       0.0782 .
## PricePerNight:RegionWashinton DC                0.5660
## UserIncome:RegionLas Vegas                      0.6446
## UserIncome:RegionMiami                          0.2685
## UserIncome:RegionWashinton DC                   0.9620
## PricePerNight:UserIncome:RegionLas Vegas        0.5586
## PricePerNight:UserIncome:RegionMiami            0.3400
## PricePerNight:UserIncome:RegionWashinton DC     0.8518
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4165 on 24984 degrees of freedom
## Multiple R-squared:  0.09847,    Adjusted R-squared:  0.09793
## F-statistic: 181.9 on 15 and 24984 DF,  p-value: < 2.2e-16
```

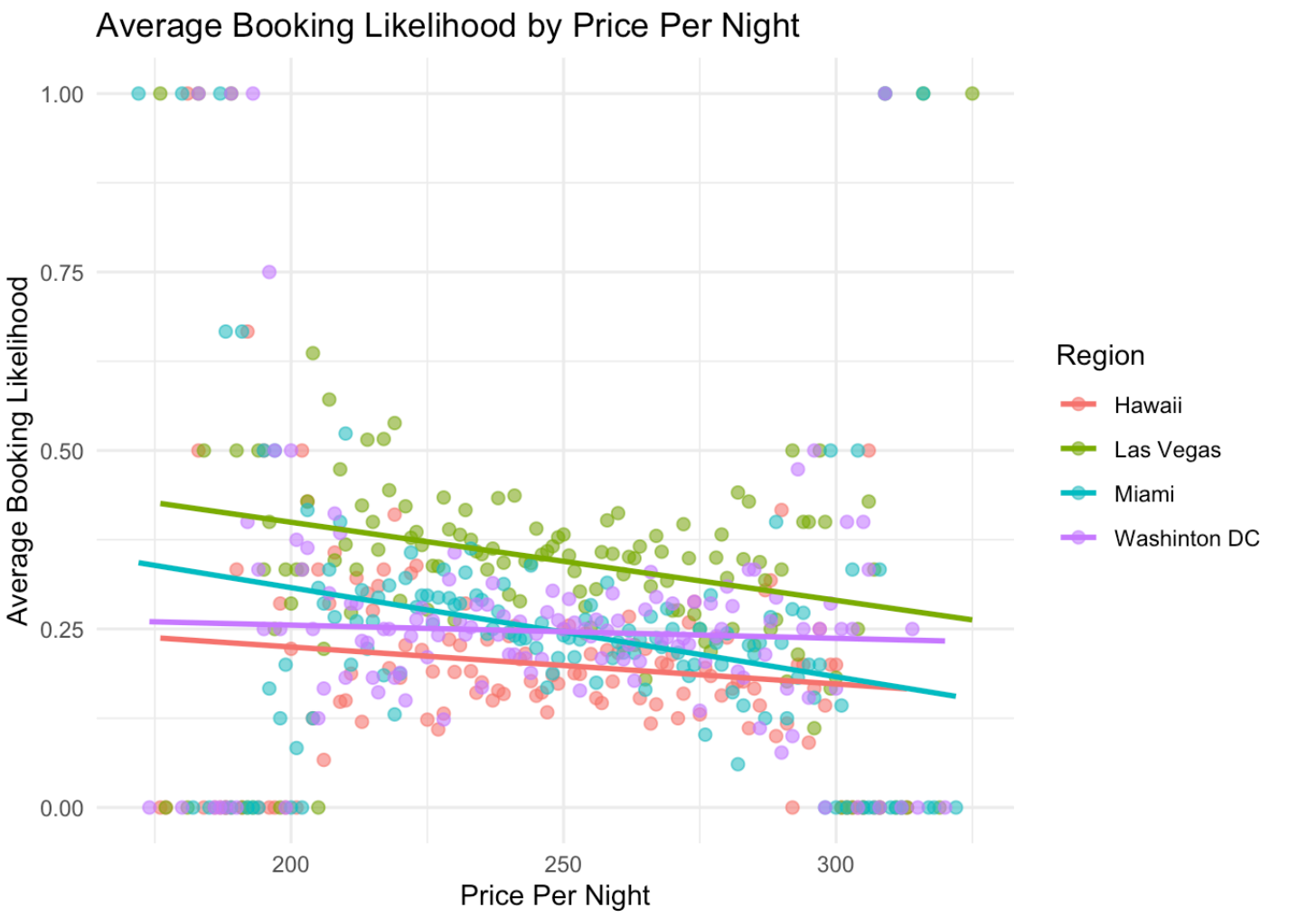
```
# Linear Regression for Number of Nights Booked with interactions
model_nights_het <- lm(Nights ~ PricePerNight * UserIncome * Region, data = booked_da
ta)
summary(model_nights_het)
```

```
##
## Call:
## lm(formula = Nights ~ PricePerNight * UserIncome * Region, data = booked_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.15734 -0.84598  0.03951  0.14483  3.03037
##
## Coefficients:
##                                               Estimate Std. Error t value
## (Intercept)                                  4.178e+00  5.427e-01   7.698
```

```
## PricePerNight                                  -4.913e-03  2.193e-03  -2.240
## UserIncome                                     -5.605e-06  6.531e-06  -0.858
## RegionLas Vegas                                -1.327e+00  6.747e-01  -1.967
## RegionMiami                                    -5.712e-01  7.165e-01  -0.797
## RegionWashinton DC                             -7.109e-01  7.006e-01  -1.015
## PricePerNight:UserIncome                        2.162e-08  2.624e-08   0.824
## PricePerNight:RegionLas Vegas                   5.497e-03  2.723e-03   2.019
## PricePerNight:RegionMiami                       2.346e-03  2.896e-03   0.810
## PricePerNight:RegionWashinton DC                2.698e-03  2.819e-03   0.957
## UserIncome:RegionLas Vegas                      1.790e-05  8.605e-06   2.080
## UserIncome:RegionMiami                          6.981e-06  8.679e-06   0.804
## UserIncome:RegionWashinton DC                   3.266e-06  8.630e-06   0.378
## PricePerNight:UserIncome:RegionLas Vegas       -7.222e-08  3.456e-08  -2.090
## PricePerNight:UserIncome:RegionMiami           -2.870e-08  3.484e-08  -0.824
## PricePerNight:UserIncome:RegionWashinton DC    -1.130e-08  3.451e-08  -0.327
##                                                Pr(>|t|)
## (Intercept)                                    1.59e-14 ***
## PricePerNight                                    0.0251 *
## UserIncome                                       0.3908
## RegionLas Vegas                                  0.0492 *
## RegionMiami                                      0.4253
## RegionWashinton DC                               0.3103
## PricePerNight:UserIncome                         0.4100
## PricePerNight:RegionLas Vegas                    0.0435 *
## PricePerNight:RegionMiami                        0.4179
## PricePerNight:RegionWashinton DC                 0.3385
## UserIncome:RegionLas Vegas                       0.0376 *
## UserIncome:RegionMiami                           0.4212
## UserIncome:RegionWashinton DC                    0.7051
## PricePerNight:UserIncome:RegionLas Vegas         0.0367 *
## PricePerNight:UserIncome:RegionMiami             0.4101
## PricePerNight:UserIncome:RegionWashinton DC      0.7434
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7547 on 6478 degrees of freedom
## Multiple R-squared:  0.006935,   Adjusted R-squared:  0.004635
## F-statistic: 3.016 on 15 and 6478 DF,  p-value: 7.314e-05
```

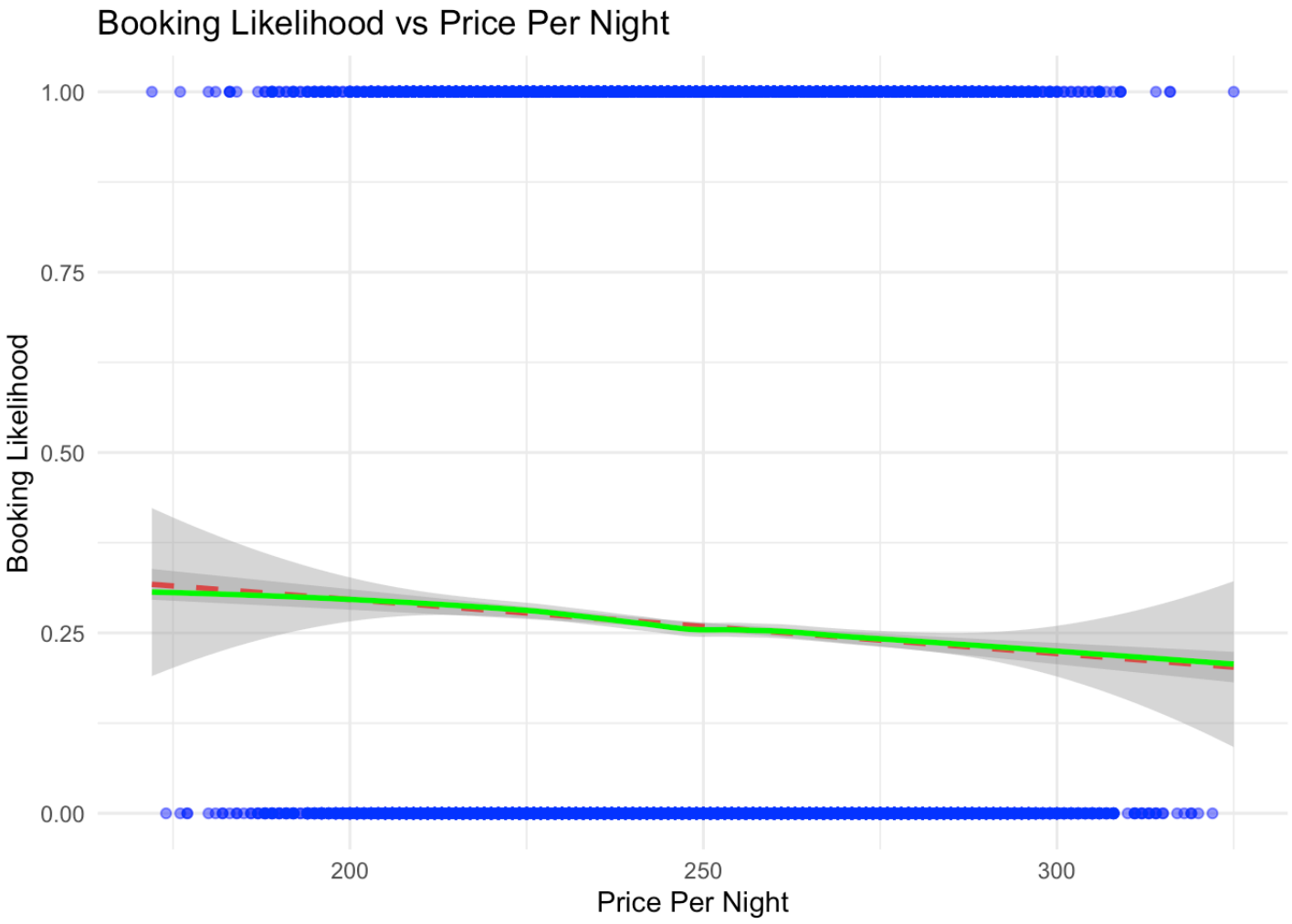
```
library(ggplot2)
library(dplyr)

# Ensure 'Booked?' is numeric
Expedia$`Booked?` <- as.numeric(Expedia$`Booked?`)

# Plot: Average Booking Likelihood by Price Per Night, color-coded by Region
ggplot(Expedia, aes(x = PricePerNight, y = `Booked?`, color = Region)) +
  stat_summary(fun = mean, geom = "point", size = 2, alpha = 0.6) +  # Plot average b
y group
  geom_smooth(method = "lm", se = FALSE) +  # Add regression lines
  labs(title = "Average Booking Likelihood by Price Per Night",
       x = "Price Per Night",
       y = "Average Booking Likelihood",
       color = "Region") +
  theme_minimal()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

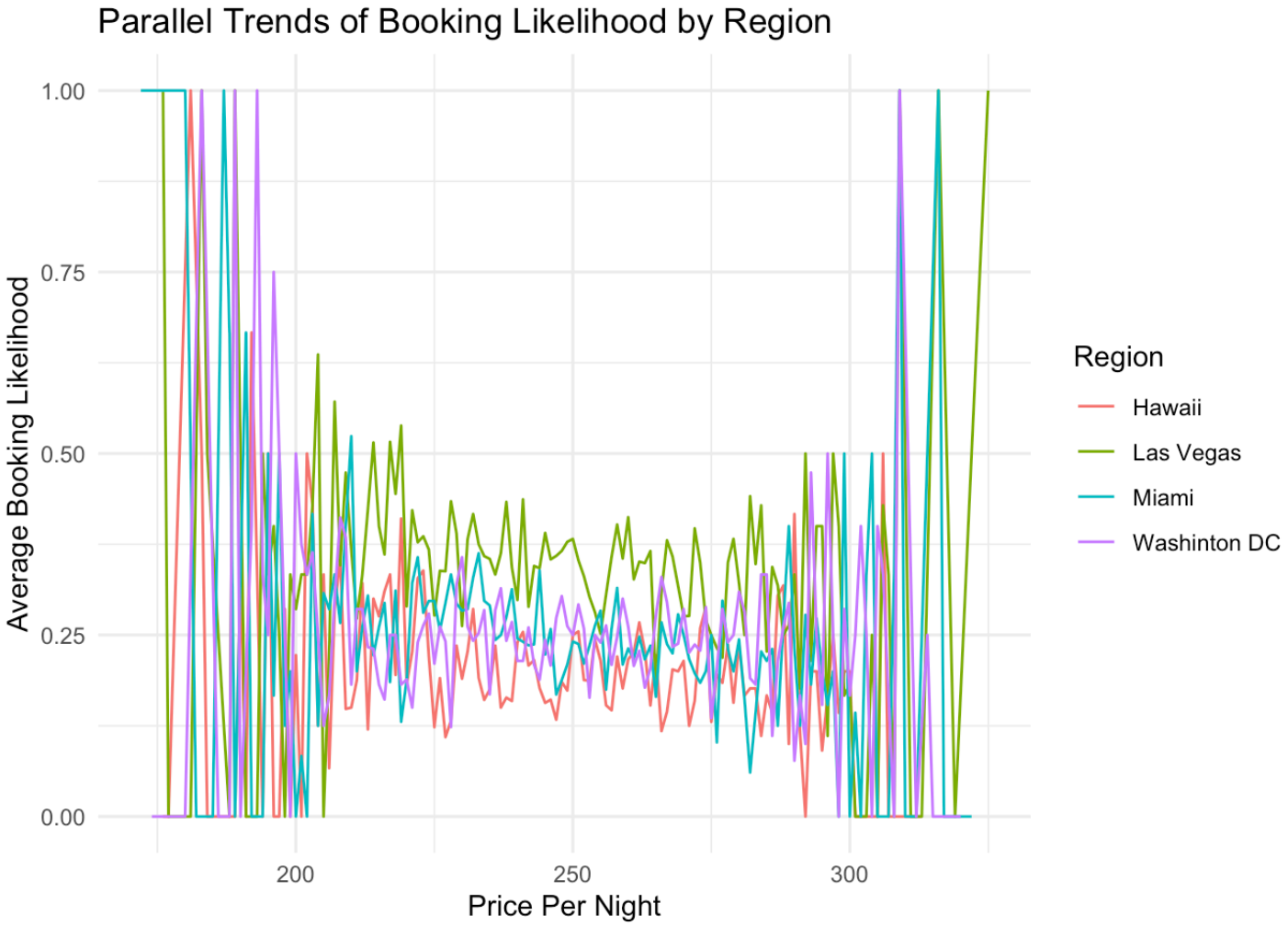


```
ggplot(Expedia, aes(x = PricePerNight, y = as.numeric(`Booked?`))) +
  geom_point(color = "blue", alpha = 0.5) +
  geom_smooth(method = "lm", color = "red", linetype = "dashed") +  # Linear trend
  geom_smooth(method = "loess", color = "green") +  # Flexible trend
  labs(title = "Booking Likelihood vs Price Per Night",
       x = "Price Per Night",
       y = "Booking Likelihood") +
  theme_minimal()
```

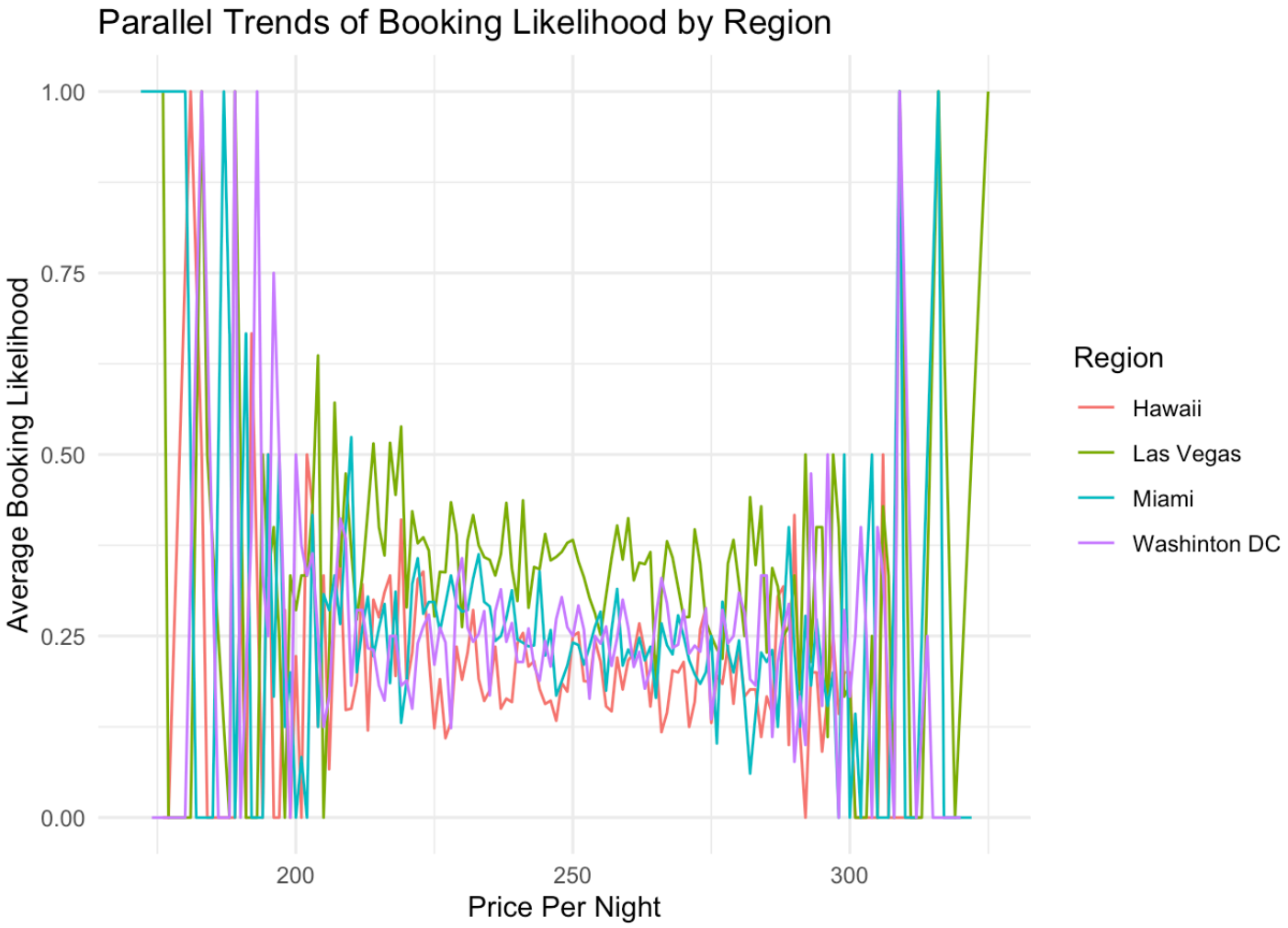
```
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
```



```
ggplot(Expedia, aes(x = PricePerNight, y = as.numeric(`Booked?`), color = Region)) +
  stat_summary(fun = mean, geom = "line") +
  labs(title = "Parallel Trends of Booking Likelihood by Region",
       x = "Price Per Night",
       y = "Average Booking Likelihood",
       color = "Region") +
  theme_minimal()
```

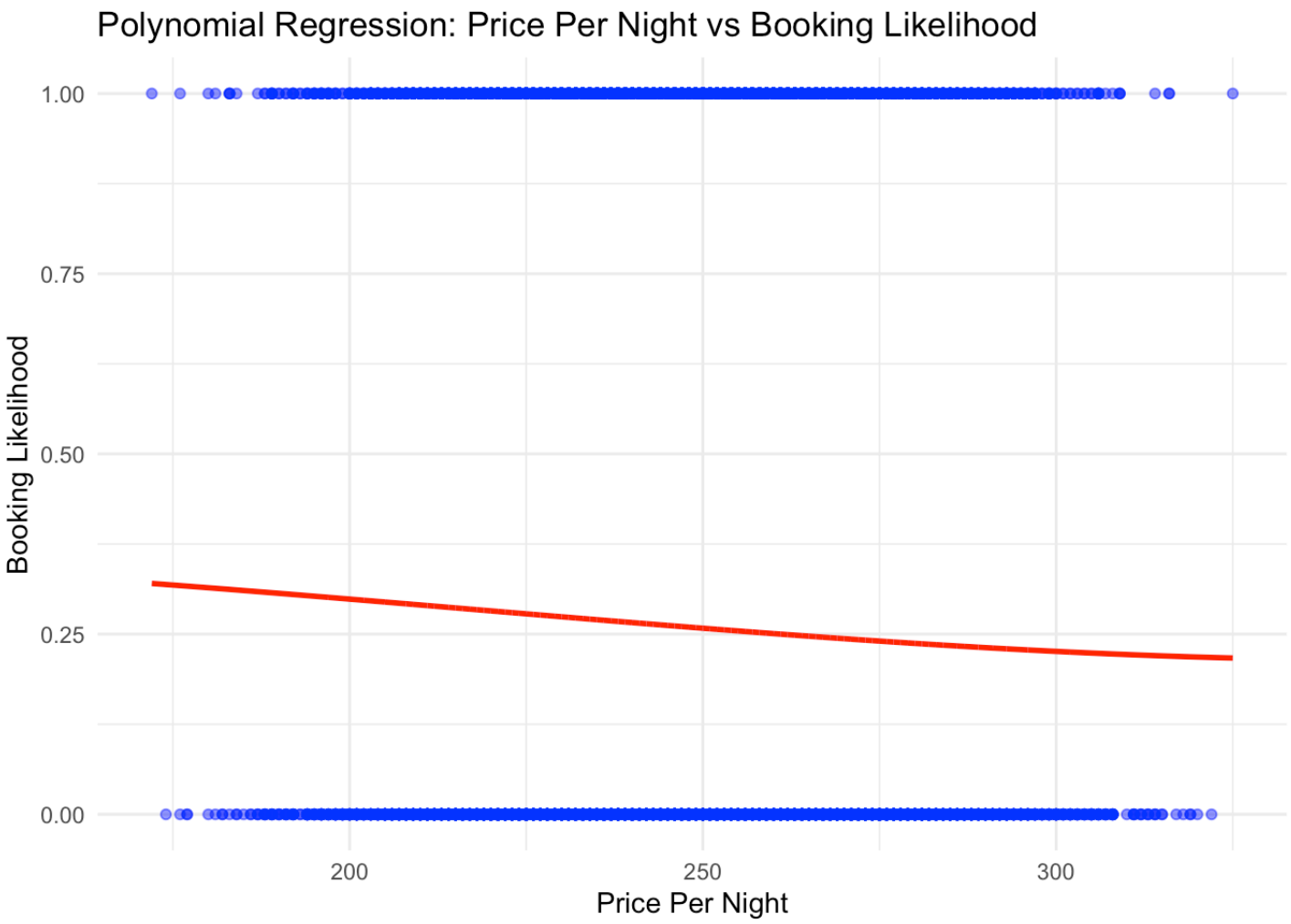



```
ggplot(Expedia, aes(x = PricePerNight, y = as.numeric(`Booked?`), color = Region)) +
  stat_summary(fun = mean, geom = "line") +
  labs(title = "Parallel Trends of Booking Likelihood by Region",
       x = "Price Per Night",
       y = "Average Booking Likelihood",
       color = "Region") +
  theme_minimal()
```



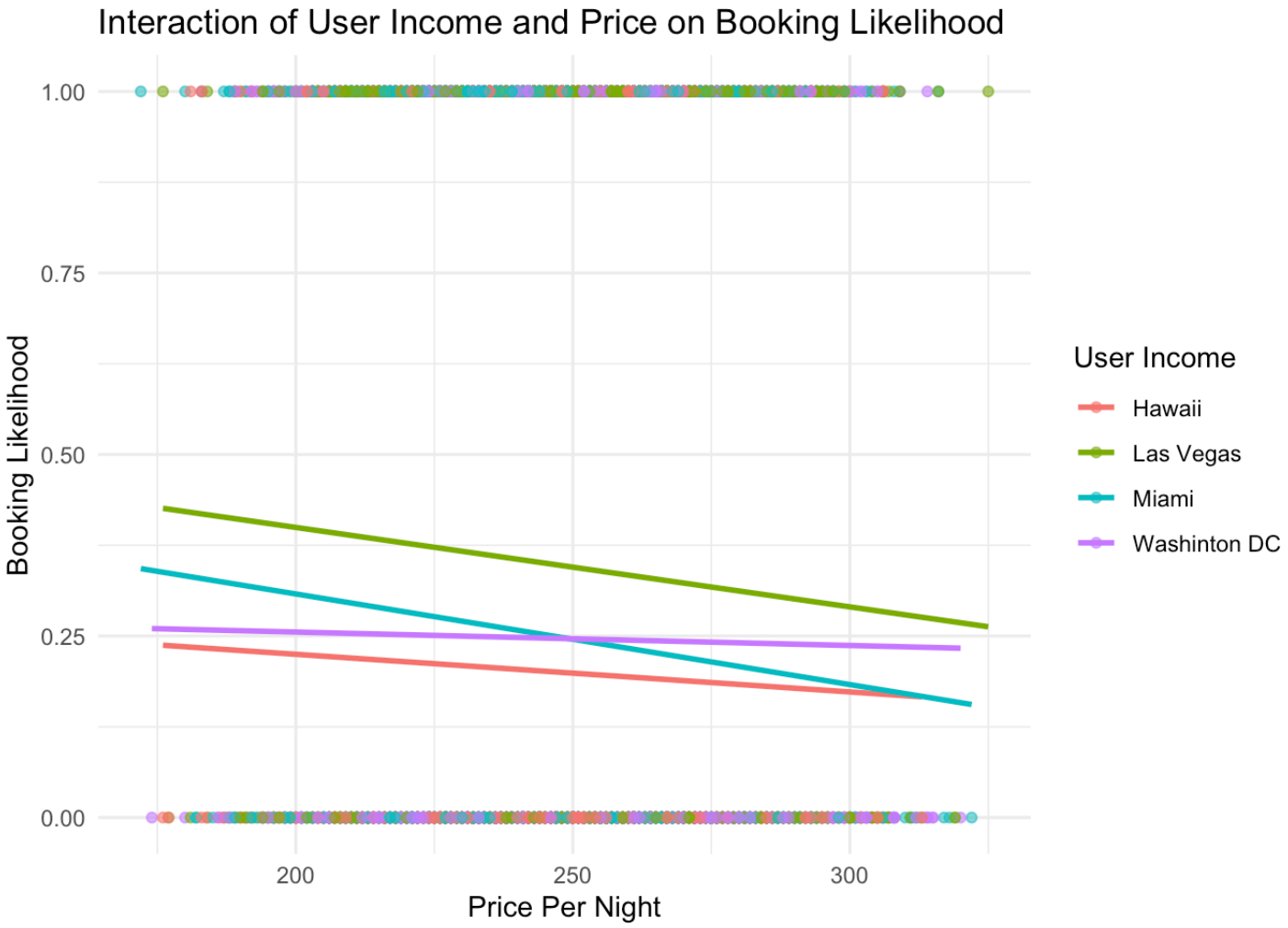

```
# Create polynomial terms
Expedia$PriceSquare <- Expedia$PricePerNight^2
Expedia$PriceCube <- Expedia$PricePerNight^3

# Fit the polynomial model
poly_model <- lm(as.numeric(`Booked?`) ~ PricePerNight + PriceSquare + PriceCube, dat
a = Expedia)
Expedia$predicted_poly <- predict(poly_model)

# Plot
ggplot(Expedia, aes(x = PricePerNight, y = as.numeric(`Booked?`))) +
  geom_point(color = "blue", alpha = 0.5) +
  geom_line(aes(y = predicted_poly), color = "red", size = 1) +
  labs(title = "Polynomial Regression: Price Per Night vs Booking Likelihood",
       x = "Price Per Night",
       y = "Booking Likelihood") +
  theme_minimal()
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## ℹ Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```




```
ggplot(Expedia, aes(x = PricePerNight, y = as.numeric(`Booked?`), color = as.factor(R
egion))) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = "Interaction of User Income and Price on Booking Likelihood",
       x = "Price Per Night",
       y = "Booking Likelihood",
       color = "User Income") +
  theme_minimal()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
ggplot(Expedia, aes(x = PricePerNight, y = as.numeric(`Booked?`), color = Region)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", se = FALSE) +  # Linear regression lines
  labs(title = "Linear Regression: Booking Likelihood by Price and Region",
       x = "Price Per Night",
       y = "Booking Likelihood",
       color = "Region") +
  theme_minimal()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```


```
ggplot(Expedia, aes(x = PricePerNight, y = as.numeric(`Booked?`))) +
  geom_point(alpha = 0.5, color = "blue") +
  geom_smooth(method = "lm", color = "red", se = FALSE) +  # Linear trend line
  labs(title = "Linear Regression: Price Per Night vs Booking Likelihood",
       x = "Price Per Night",
       y = "Booking Likelihood") +
  theme_minimal()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```


```
# Define custom income categories
Expedia$IncomeCategory <- cut(Expedia$UserIncome,
                              breaks = c(0, 50000, 100000, Inf),
                              labels = c("0-50k", "50-100k", "100k+"),
                              include.lowest = TRUE)

# Plot with categorized income groups
ggplot(Expedia, aes(x = PricePerNight, y = as.numeric(`Booked?`), color = IncomeCateg
ory)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", se = FALSE) +  # Linear regression lines by income categ
ory
  labs(title = "Linear Regression: Interaction of Income Category and Price",
       x = "Price Per Night",
       y = "Booking Likelihood",
       color = "Income Category") +
  theme_minimal()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```


```
# Filter the data for booked cases only
booked_data <- Expedia %>% filter(`Booked?` == 1)

# Plot for number of nights booked
ggplot(booked_data, aes(x = PricePerNight, y = Nights)) +
  geom_point(alpha = 0.5, color = "green") +
  geom_smooth(method = "lm", color = "red", se = FALSE) +  # Linear trend line
  labs(title = "Linear Regression: Price Per Night vs Number of Nights Booked",
       x = "Price Per Night",
       y = "Number of Nights Booked") +
  theme_minimal()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```


```
# final bucket of user income
# Grouping UserIncome into specified buckets
Expedia$IncomeCategory <- cut(Expedia$UserIncome,
                              breaks = c(0, 50000, 100000, Inf),
                              labels = c("0-50k", "50-100k", "100k+"),
                              include.lowest = TRUE)

# Convert IncomeCategory to factor
Expedia$IncomeCategory <- factor(Expedia$IncomeCategory)
```

```
# Final 2
# Linear regression for Booking Likelihood (PricePerNight, IncomeCategory, Region)
model_booking_het <- lm(as.numeric(`Booked?`) ~ PricePerNight * IncomeCategory * Regi
on, data = Expedia)

# Display summary of the model
summary(model_booking_het)
```

```
##
## Call:
## lm(formula = as.numeric(`Booked?`) ~ PricePerNight * IncomeCategory *
##     Region, data = Expedia)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.7250 -0.2736 -0.1667  0.3617  0.9252
##
## Coefficients:
##                                                    Estimate Std. Error
## (Intercept)                                       2.552e-01  8.556e-02
## PricePerNight                                    -5.407e-04  3.432e-04
## IncomeCategory50-100k                             2.659e-01  1.401e-01
## IncomeCategory100k+                               3.271e-02  2.626e-01
## RegionLas Vegas                                   3.039e-01  1.207e-01
## RegionMiami                                       2.730e-01  1.220e-01
## RegionWashinton DC                                2.935e-02  1.207e-01
## PricePerNight:IncomeCategory50-100k              -5.028e-04  5.621e-04
## PricePerNight:IncomeCategory100k+                 1.663e-03  1.048e-03
## PricePerNight:RegionLas Vegas                    -6.174e-04  4.837e-04
## PricePerNight:RegionMiami                        -8.943e-04  4.887e-04
## PricePerNight:RegionWashinton DC                  7.244e-05  4.838e-04
## IncomeCategory50-100k:RegionLas Vegas            -5.055e-02  1.992e-01
## IncomeCategory100k+:RegionLas Vegas              -2.418e-01  3.724e-01
## IncomeCategory50-100k:RegionMiami                -6.152e-02  1.992e-01
## IncomeCategory100k+:RegionMiami                  -3.720e-01  3.612e-01
## IncomeCategory50-100k:RegionWashinton DC         -1.662e-01  1.996e-01
## IncomeCategory100k+:RegionWashinton DC           -1.768e-02  3.651e-01
## PricePerNight:IncomeCategory50-100k:RegionLas Vegas  2.104e-04  7.981e-04
## PricePerNight:IncomeCategory100k+:RegionLas Vegas    7.537e-04  1.490e-03
## PricePerNight:IncomeCategory50-100k:RegionMiami      2.247e-04  7.987e-04
## PricePerNight:IncomeCategory100k+:RegionMiami        1.311e-03  1.441e-03
## PricePerNight:IncomeCategory50-100k:RegionWashinton DC  6.780e-04  7.998e-04
## PricePerNight:IncomeCategory100k+:RegionWashinton DC   -9.871e-05  1.455e-03
##                                                   t value Pr(>|t|)
## (Intercept)                                         2.983  0.00286 **
## PricePerNight                                      -1.575  0.11517
```

```
## IncomeCategory50-100k                                    1.898  0.05777 .
## IncomeCategory100k+                                      0.125  0.90090
## RegionLas Vegas                                          2.519  0.01179 *
## RegionMiami                                              2.239  0.02517 *
## RegionWashinton DC                                       0.243  0.80792
## PricePerNight:IncomeCategory50-100k                     -0.895  0.37104
## PricePerNight:IncomeCategory100k+                        1.586  0.11268
## PricePerNight:RegionLas Vegas                           -1.277  0.20178
## PricePerNight:RegionMiami                               -1.830  0.06726 .
## PricePerNight:RegionWashinton DC                         0.150  0.88096
## IncomeCategory50-100k:RegionLas Vegas                   -0.254  0.79967
## IncomeCategory100k+:RegionLas Vegas                     -0.649  0.51617
## IncomeCategory50-100k:RegionMiami                       -0.309  0.75747
## IncomeCategory100k+:RegionMiami                         -1.030  0.30299
## IncomeCategory50-100k:RegionWashinton DC                -0.833  0.40502
## IncomeCategory100k+:RegionWashinton DC                  -0.048  0.96137
## PricePerNight:IncomeCategory50-100k:RegionLas Vegas      0.264  0.79211
## PricePerNight:IncomeCategory100k+:RegionLas Vegas        0.506  0.61292
## PricePerNight:IncomeCategory50-100k:RegionMiami          0.281  0.77846
## PricePerNight:IncomeCategory100k+:RegionMiami            0.910  0.36269
## PricePerNight:IncomeCategory50-100k:RegionWashinton DC   0.848  0.39665
## PricePerNight:IncomeCategory100k+:RegionWashinton DC    -0.068  0.94592
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4199 on 24976 degrees of freedom
## Multiple R-squared:  0.08403,    Adjusted R-squared:  0.08319
## F-statistic: 99.63 on 23 and 24976 DF,  p-value: < 2.2e-16
```

```
# Extract the coefficient for PricePerNight
price_coef <- coef(model_booking_het)["PricePerNight"]

# Calculate the impact of a $100 increase
impact_100_dollars <- price_coef * 100
print(paste("Estimated impact of a $100 increase: ", impact_100_dollars))
```

```
## [1] "Estimated impact of a $100 increase:  -0.0540666406503923"
```

```
#FINAL 2
# Filter for users who have booked
booked_data <- Expedia %>% filter(`Booked?` == 1)

# Run linear regression for Number of Nights Booked
model_nights_het <- lm(Nights ~ PricePerNight * IncomeCategory * Region, data = booke
d_data)

# Display summary of the model
summary(model_nights_het)
```

```
##
## Call:
## lm(formula = Nights ~ PricePerNight * IncomeCategory * Region,
##     data = booked_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.20859 -0.83881  0.03424  0.14630  3.02381
##
## Coefficients:
##                                                    Estimate Std. Error
## (Intercept)                                       3.7778922  0.4538405
## PricePerNight                                    -0.0033010  0.0018343
## IncomeCategory50-100k                             0.3624510  0.5885934
## IncomeCategory100k+                              -0.9526172  0.7549493
## RegionLas Vegas                                  -0.6083719  0.5418754
## RegionMiami                                       0.2156238  0.5986011
## RegionWashinton DC                               -0.2060816  0.5772392
## PricePerNight:IncomeCategory50-100k              -0.0015727  0.0023777
## PricePerNight:IncomeCategory100k+                 0.0037069  0.0030193
## PricePerNight:RegionLas Vegas                     0.0025503  0.0021888
## PricePerNight:RegionMiami                        -0.0008795  0.0024236
## PricePerNight:RegionWashinton DC                  0.0008505  0.0023284
## IncomeCategory50-100k:RegionLas Vegas             0.4868098  0.7313623
## IncomeCategory100k+:RegionLas Vegas               1.7035769  0.9869409
## IncomeCategory50-100k:RegionMiami                -1.0931839  0.7962038
## IncomeCategory100k+:RegionMiami                   1.0570131  1.0038838
## IncomeCategory50-100k:RegionWashinton DC         -0.9516053  0.7777092
## IncomeCategory100k+:RegionWashinton DC            0.8205010  1.0145967
## PricePerNight:IncomeCategory50-100k:RegionLas Vegas -0.0018637  0.0029517
## PricePerNight:IncomeCategory100k+:RegionLas Vegas -0.0068329  0.0039500
## PricePerNight:IncomeCategory50-100k:RegionMiami    0.0044560  0.0032213
## PricePerNight:IncomeCategory100k+:RegionMiami     -0.0042568  0.0040139
## PricePerNight:IncomeCategory50-100k:RegionWashinton DC  0.0036338  0.0031331
## PricePerNight:IncomeCategory100k+:RegionWashinton DC   -0.0031380  0.0040496
```

```
##                                                    t value Pr(>|t|)
## (Intercept)                                          8.324   <2e-16 ***
## PricePerNight                                       -1.800   0.0720 .
## IncomeCategory50-100k                                0.616   0.5381
## IncomeCategory100k+                                 -1.262   0.2071
## RegionLas Vegas                                     -1.123   0.2616
## RegionMiami                                          0.360   0.7187
## RegionWashinton DC                                  -0.357   0.7211
## PricePerNight:IncomeCategory50-100k                 -0.661   0.5084
## PricePerNight:IncomeCategory100k+                    1.228   0.2196
## PricePerNight:RegionLas Vegas                        1.165   0.2440
## PricePerNight:RegionMiami                           -0.363   0.7167
## PricePerNight:RegionWashinton DC                     0.365   0.7149
## IncomeCategory50-100k:RegionLas Vegas                0.666   0.5057
## IncomeCategory100k+:RegionLas Vegas                  1.726   0.0844 .
## IncomeCategory50-100k:RegionMiami                   -1.373   0.1698
## IncomeCategory100k+:RegionMiami                      1.053   0.2924
## IncomeCategory50-100k:RegionWashinton DC            -1.224   0.2211
## IncomeCategory100k+:RegionWashinton DC               0.809   0.4187
## PricePerNight:IncomeCategory50-100k:RegionLas Vegas -0.631   0.5278
## PricePerNight:IncomeCategory100k+:RegionLas Vegas   -1.730   0.0837 .
## PricePerNight:IncomeCategory50-100k:RegionMiami      1.383   0.1666
## PricePerNight:IncomeCategory100k+:RegionMiami       -1.061   0.2889
## PricePerNight:IncomeCategory50-100k:RegionWashinton DC   1.160   0.2462
## PricePerNight:IncomeCategory100k+:RegionWashinton DC    -0.775   0.4384
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7546 on 6470 degrees of freedom
## Multiple R-squared:  0.008321,   Adjusted R-squared:  0.004795
## F-statistic:  2.36 on 23 and 6470 DF,  p-value: 0.0002555
```

```
# Extract the coefficient for PricePerNight
price_coef_nights <- coef(model_nights_het)["PricePerNight"]

# Calculate the impact of a $100 increase on number of nights
impact_100_dollars_nights <- price_coef_nights * 100
print(paste("Estimated impact of a $100 increase on nights booked: ", impact_100_doll
ars_nights))
```

```
## [1] "Estimated impact of a $100 increase on nights booked:  -0.33009759470085"
```