

```
In [1]: # 3.2 Assignment Using Data to Improve a Marketing Promotion
# Manuel Duran DSC 630
```

```
In [2]: # Importation of Library and Data File
```

```
In [4]: library("ggplot2")
library('magrittr')
```

```
In [5]: dodgers <- read.csv(file = 'dodgers.csv')
```

```
In [6]: # Data Review
head(dodgers)
```

month	day	attend	day_of_week	opponent	temp	skies	day_night	cap	shirt	fireworks	bobbleh
APR	10	56000	Tuesday	Pirates	67	Clear	Day	NO	NO	NO	
APR	11	29729	Wednesday	Pirates	58	Cloudy	Night	NO	NO	NO	
APR	12	28328	Thursday	Pirates	57	Cloudy	Night	NO	NO	NO	
APR	13	31601	Friday	Padres	54	Cloudy	Night	NO	NO	YES	
APR	14	46549	Saturday	Padres	57	Cloudy	Night	NO	NO	NO	
APR	15	38359	Sunday	Padres	65	Clear	Day	NO	NO	NO	

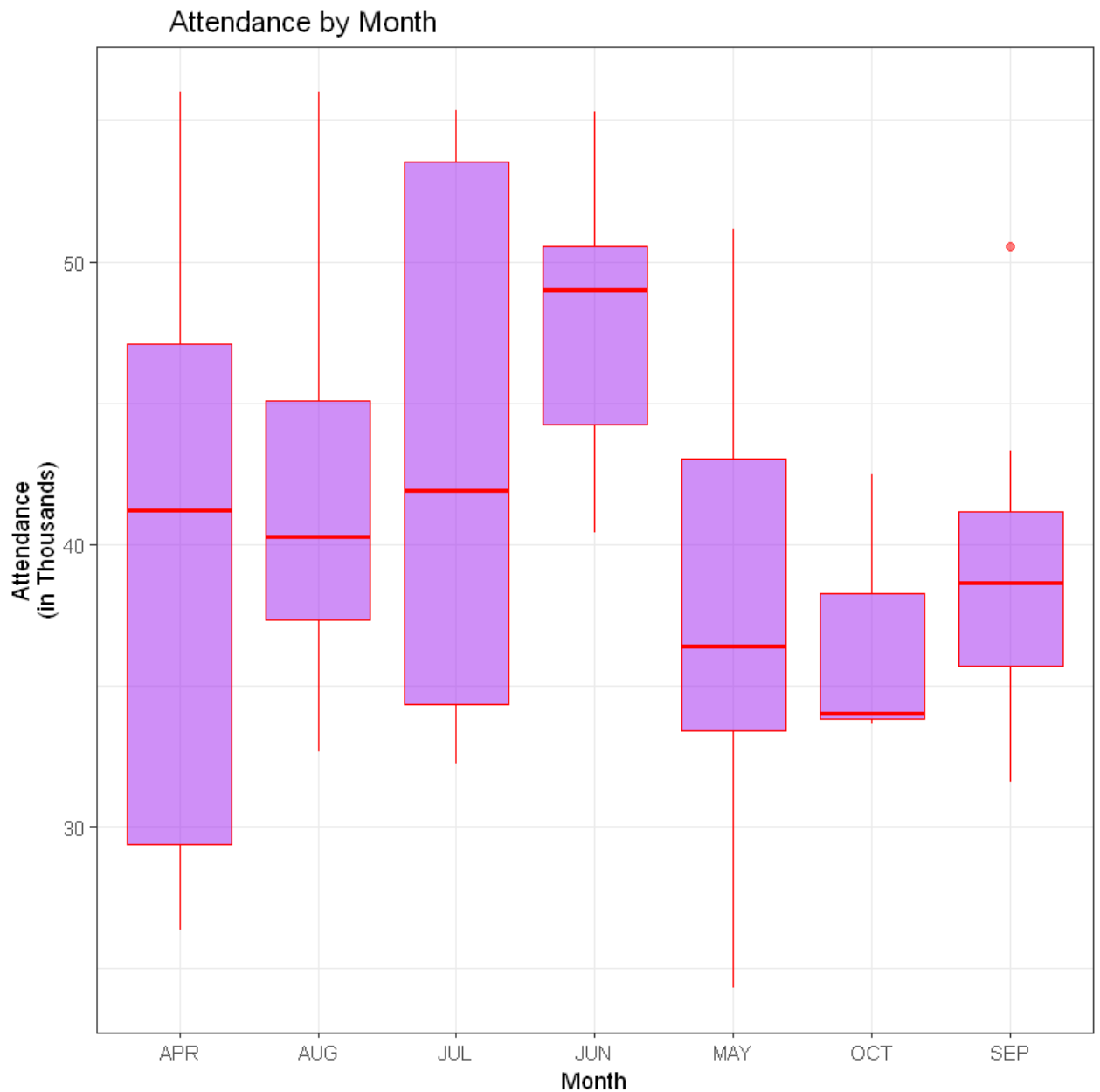
```
In [7]: # Creation of BoxPlots
```

```
In [8]: # Attendance by day of Month
```

```
In [9]: ggplot2::ggplot(dodgers, ggplot2::aes(x=month, y=attend/1000)) +

ggplot2::geom_boxplot(color="red", fill="purple", alpha=0.5) +
  ggplot2::labs(title='Attendance by Month',
                x='Month',
                y='Attendance\n(in Thousands)') +

ggplot2::theme_bw() +
  ggplot2::theme(plot.title = ggplot2::element_text(hjust = 0.10))
```

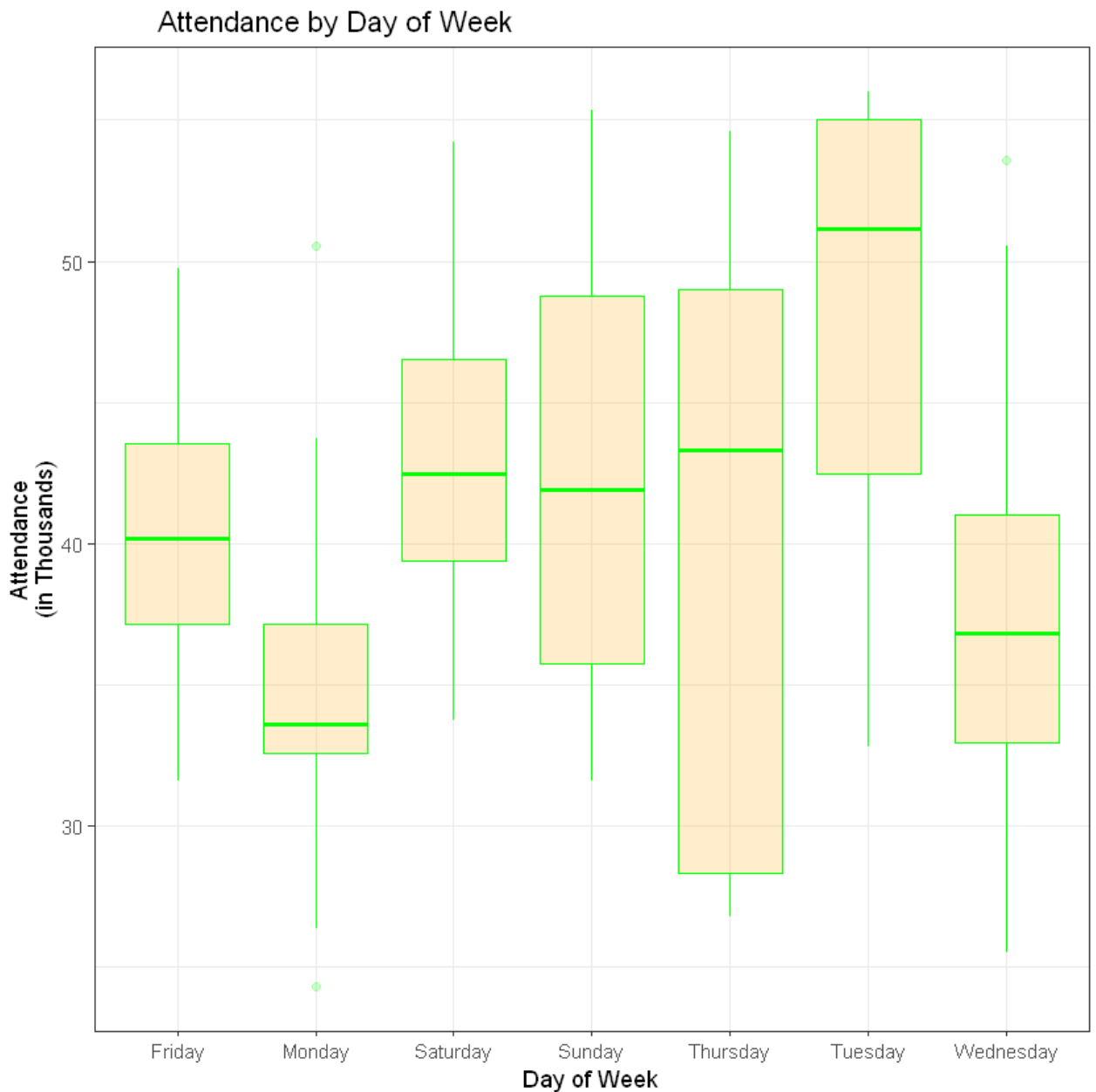


```
In [10]: # Boxplot of Attendance by Day of the of Week
```

```
In [11]: ggplot2::ggplot(dodgers, ggplot2::aes(x=day_of_week, y=attend/1000)) +

ggplot2::geom_boxplot(color="green", fill="orange", alpha=0.2) +
  ggplot2::labs(title='Attendance by Day of Week',
    x='Day of Week',
    y='Attendance\n(in Thousands)') +

ggplot2::theme_bw() +
  ggplot2::theme(plot.title = ggplot2::element_text(hjust = 0.10))
```



### Discoveries from BoxPlots

\*Attendance by Month: By Viewing the Boxplot I am beginning to suggest a promo in the months of Jun, July or August

\*Attendance by DOW: By Viewing the Boxplot I am beginning to suggest a promo on either Tues. or Thurs.

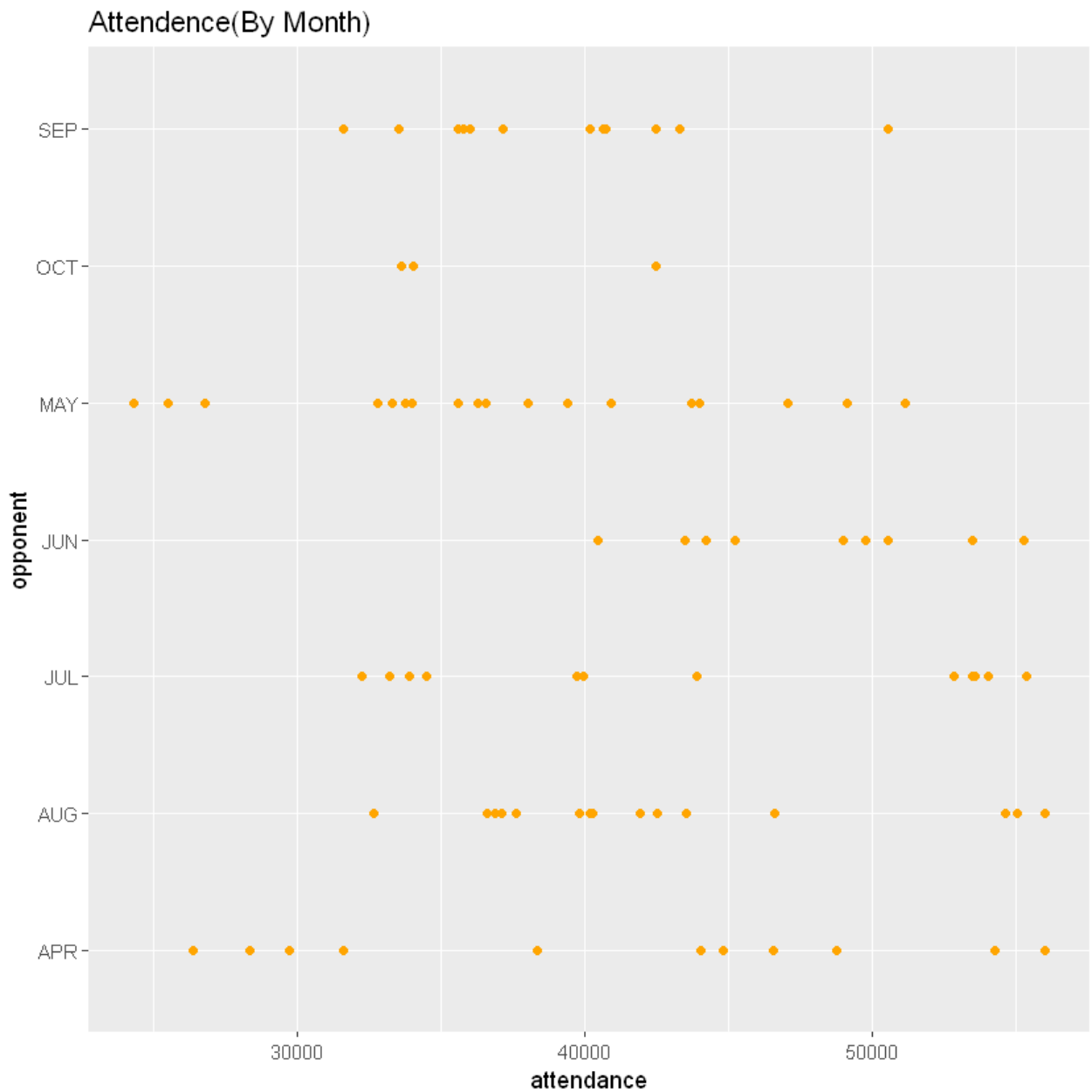
```
In [19]: # Creation of Scatter Plots
```

```
In [18]: # Attendance By Month Scatter Plot

ggplot2::ggplot(dodgers, ggplot2::aes(x=attend, y=month)) +

  ggplot2::geom_point(color="orange") +
```

```
ggplot2::labs(x='attendance',
              y='opponent',
              title = 'Attendance(By Month)', color='blue')
```



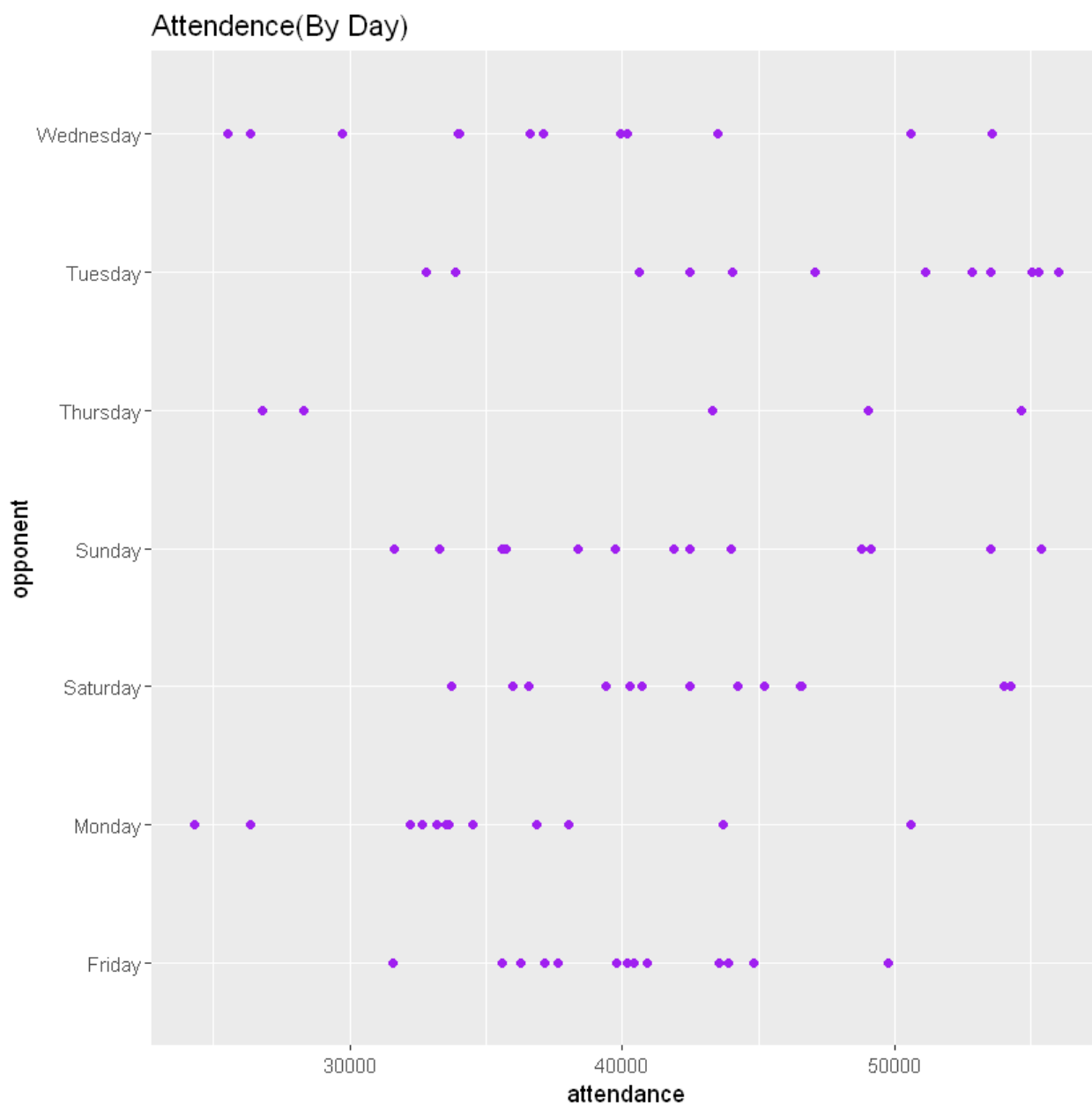
In [17]:

```
# Attendance By Day Scatter Plot

ggplot2::ggplot(dodgers, ggplot2::aes(x=attend, y=day_of_week)) +

  ggplot2::geom_point(color="purple") +

ggplot2::labs(x='attendance',
              y='opponent',
              title = 'Attendance(By Day)', color='blue')
```

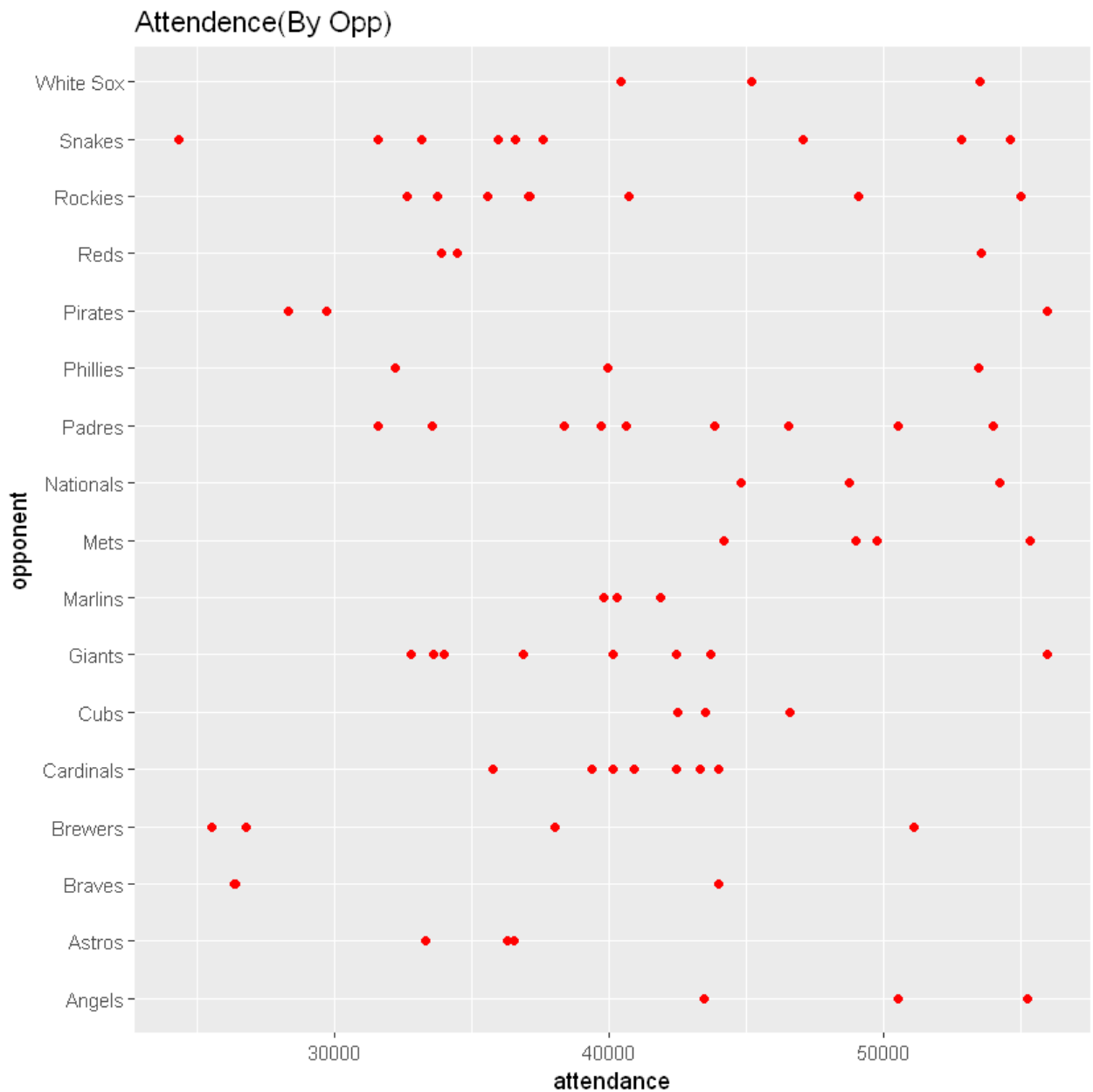


In [69]:

```
# # Attendance By opponent Scatter Plot
ggplot2::ggplot(dodgers, ggplot2::aes(x=attend, y=opponent)) +

  ggplot2::geom_point(color="red") +

ggplot2::labs(x='attendance',
              y='opponent',
              title = 'Attendance(By Opp)', color='blue')
```



### Discoveries from Scatter Plots

\*Attendance (by Month): July and August produce the strongest attendance months with more days nearing 50K in attendance.

\*Attendance (by Day): Tuesday Saturday and Sunday Produce the strongest attence numbers.

\*This is due to more people being off on weekends and many work promotions for ballparks are produced on Tuesdays

\*Attendance (by Opp): Giants, Angels, Padres, and Mets Produce the highest attendance numbers.

\*This is due to NL West Rivarlies and Higher Populated areas.

\*Best Day For Promotion Before Model Creation: Tuesday Night In August when Dogders are taking on the Giants.

```
In [13]: # Creation of Regression Model
```

```
In [20]: # Model Using Fireworks and Bobblehead Promotions
# Firework Promotion Model
```

```
In [49]: # Input Model
fw.model <- {attend ~ month + day_of_week + bobblehead}
set.seed(1234)
t_test <- c(rep(1, trunc((2/3)*nrow(dodgers))), rep(2, trunc((1/3)*nrow(dodgers))))
```

```
In [51]: # Devleopment of training and testing sets
dodgers$t_test <- sample(t_test)
dodgers$t_test <- factor(dodgers$t_test, levels = c(1, 2), labels = c("TRAIN", "TEST"))
dodgers.Tr <- subset(dodgers, t_test == "TRAIN")
dodgers.Te <- subset(dodgers, t_test == "TEST")
```

```
In [52]: # Fitting, Prediction, and Evauluation
t.model.fit <- lm(my.model, data = dodgers.Tr)
dodgers.Tr$Predict_Attend <- predict(t.model.fit)
dodgers.Te$Predict_Attend <- predict(t.model.fit, newdata = dodgers.Te)
```

```
In [53]: cat("\n", "Var accounted for: ", round(cor(dodgers.Te$attend, dodgers.Te$Predict_Attend)
```

Var accounted for: 0.103

```
In [54]: fw.model.fit <- lm(fw.model, data = dodgers) # use all available data
print(summary(fw.model.fit))
```

Call:

lm(formula = fw.model, data = dodgers)

Residuals:

Min	1Q	Median	3Q	Max
-10786.5	-3628.1	-516.1	2230.2	14351.0

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	38792.98	2364.68	16.405	< 2e-16 ***
monthAUG	2377.92	2402.91	0.990	0.3259
monthJUL	2849.83	2578.60	1.105	0.2730
monthJUN	7163.23	2732.72	2.621	0.0108 *
monthMAY	-2385.62	2291.22	-1.041	0.3015
monthOCT	-662.67	4046.45	-0.164	0.8704
monthSEP	29.03	2521.25	0.012	0.9908
day_of_weekMonday	-4883.82	2504.65	-1.950	0.0554 .
day_of_weekSaturday	1488.24	2442.68	0.609	0.5444
day_of_weekSunday	1840.18	2426.79	0.758	0.4509
day_of_weekThursday	-4108.45	3381.22	-1.215	0.2286
day_of_weekTuesday	3027.68	2686.43	1.127	0.2638
day_of_weekWednesday	-2423.80	2485.46	-0.975	0.3330
bobbleheadYES	10714.90	2419.52	4.429	3.59e-05 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6120 on 67 degrees of freedom  
 Multiple R-squared: 0.5444, Adjusted R-squared: 0.456  
 F-statistic: 6.158 on 13 and 67 DF, p-value: 2.083e-07

In [21]: `# Bobblehead Model`

In [39]: `# Input Model`  
`my.model <- {attend ~ month + day_of_week + bobblehead}`  
`set.seed(1234)`  
`training_test <- c(rep(1, trunc((2/3)*nrow(dodgers))), rep(2, trunc((1/3)*nrow(dodgers)))`

In [44]: `# Devleopment of training and testing sets`  
`dodgers$Training_Test <- sample(training_test)`  
`dodgers$Training_Test <- factor(dodgers$Training_Test, levels = c(1, 2), labels = c("TR`  
`dodgers.Train <- subset(dodgers, Training_Test == "TRAIN")`  
`dodgers.Test <- subset(dodgers, Training_Test == "TEST")`

In [47]: `# Fitting, Prediction, and Evauluation`  
`train.model.fit <- lm(my.model, data = dodgers.Train)`  
`dodgers.Train$Predict_Attend <- predict(train.model.fit)`  
`dodgers.Test$Predict_Attend <- predict(train.model.fit, newdata = dodgers.Test)`

In [42]: `cat("\n", "Var accounted for: ", round(cor(dodgers.Test$attend, dodgers.Test$Predict_Att`  
 Var account for: 0.383

In [43]: `my.model.fit <- lm(my.model, data = dodgers) # use all available data`  
`print(summary(my.model.fit))`

Call:

`lm(formula = my.model, data = dodgers)`

Residuals:

	Min	1Q	Median	3Q	Max
	-10786.5	-3628.1	-516.1	2230.2	14351.0

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	38792.98	2364.68	16.405	< 2e-16 ***
monthAUG	2377.92	2402.91	0.990	0.3259
monthJUL	2849.83	2578.60	1.105	0.2730
monthJUN	7163.23	2732.72	2.621	0.0108 *
monthMAY	-2385.62	2291.22	-1.041	0.3015
monthOCT	-662.67	4046.45	-0.164	0.8704
monthSEP	29.03	2521.25	0.012	0.9908
day_of_weekMonday	-4883.82	2504.65	-1.950	0.0554 .
day_of_weekSaturday	1488.24	2442.68	0.609	0.5444
day_of_weekSunday	1840.18	2426.79	0.758	0.4509
day_of_weekThursday	-4108.45	3381.22	-1.215	0.2286
day_of_weekTuesday	3027.68	2686.43	1.127	0.2638
day_of_weekWednesday	-2423.80	2485.46	-0.975	0.3330
bobbleheadYES	10714.90	2419.52	4.429	3.59e-05 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1



Residual standard error: 6120 on 67 degrees of freedom  
Multiple R-squared: 0.5444, Adjusted R-squared: 0.456  
F-statistic: 6.158 on 13 and 67 DF, p-value: 2.083e-07

In [ ]:

```
# Final Conclusions of Assignment
```

### Discoveries from Regession Analysis

*From my regression models we can see a direct impact from a promo being added in. For my selection of when the best time to run a marketing promo is based on he regression model: Would be the month of June on a Tuesday Night. Both T-Values within the model for June and Tuesday have great predictive power as coeffcients. \*Scatterplots and Boxplots have showed a pattern leading to this conclusion and the model has proven the idea as true*