

512 Project Code

null

Loading data

```
df <- read.csv("PS_2023.10.28_11.10.58.csv")
```

Removing planets that have a controversial flag (whether the confirmation status of a planet has been questioned in the published literature)

```
filtered_df <- df %>% filter(pl_controv_flag != 1)
```

Remove duplicates based on lowest average distance error and keep the most recent publication date in case of ties

```
filtered_df$PublicationDate <- as.Date(paste0(filtered_df$pl_pubdate, "-01"))
filtered_df <- filtered_df %>%
  mutate(AvgDistError = (sy_disterr1 + sy_disterr1) / 2) %>%
  arrange(pl_name, AvgDistError, PublicationDate) %>%
  group_by(pl_name) %>%
  filter(AvgDistError == min(AvgDistError)) %>%
  slice_tail(n = 1) %>%
  ungroup()
filtered_df <- filtered_df %>%
  arrange(pl_name) %>%
  group_by(hostname) %>%
  slice_head(n=1) %>%
  ungroup()
```

Only keeping the columns that we are interested in

```
compact_df <- filtered_df %>% select(pl_name, hostname, discoverymethod, pl_orbper, pl_orbsmax, pl_rade, pl_b
```

Renaming columns to more clear names

```
compact_df <- compact_df %>% dplyr::rename(OrbitalPeriod = pl_orbper, SemiMajorAxis = pl_orbsmax, Radius
```

Visualizing missing data

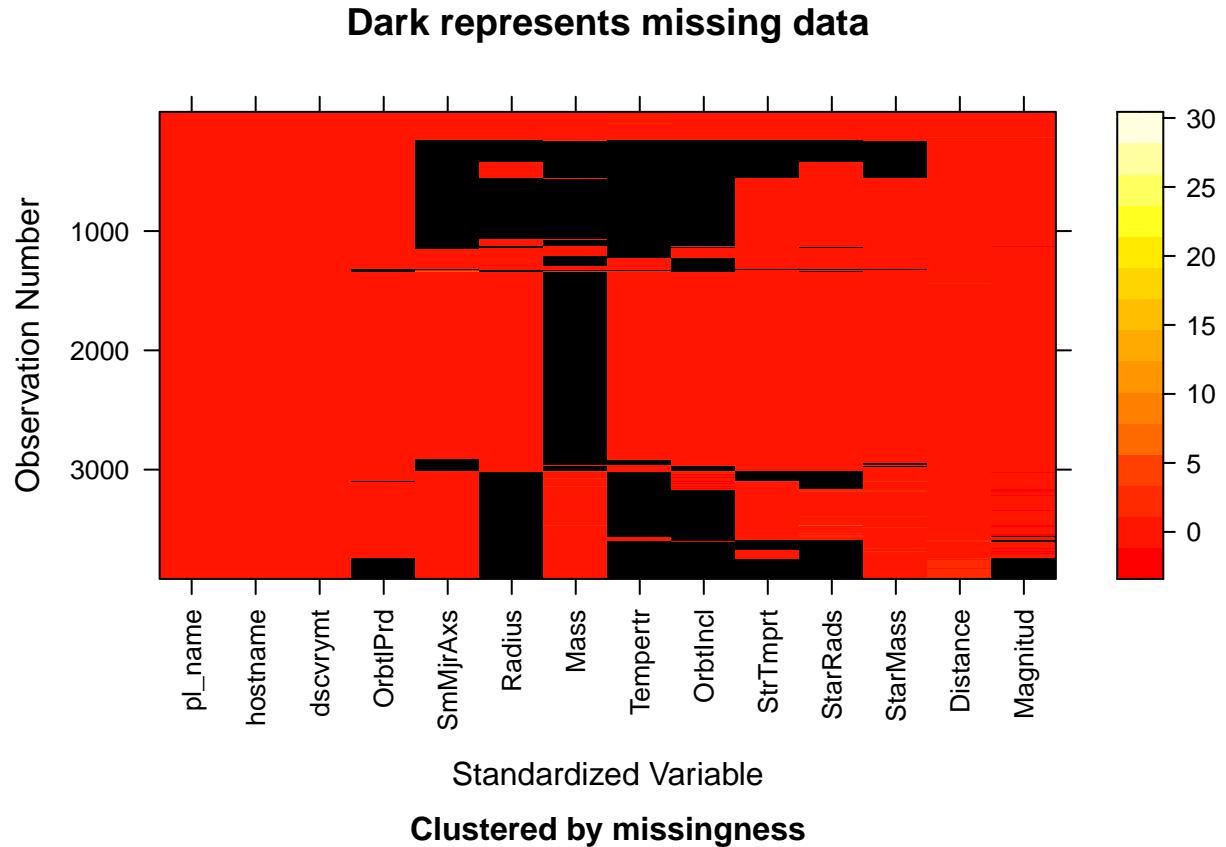
```
#compact_df <- compact_df %>% filter(StarSpecType!="")
library(mi)
mdf <- missing_data.frame(as.data.frame(compact_df))
```

```

## NOTE: In the following pairs of variables, the missingness pattern of the second is a subset of the first.
## Please verify whether they are in fact logically distinct variables.
##      [,1]      [,2]
## [1,] "Radius" "Magnitude"

```

```
image(mdf)
```



Clustered by missingness

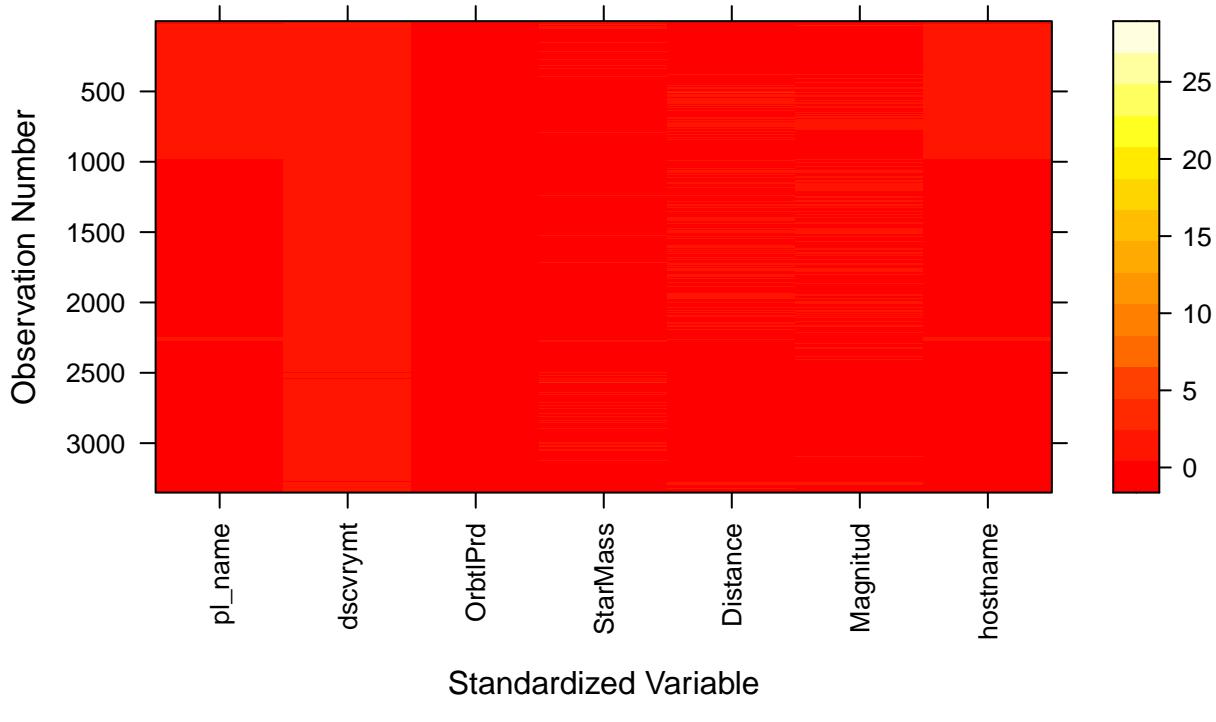
Based on the missing data visualization and for the sake of simplicity of the analysis, I will only keep the following variables: planet name(for identification only), discovery method, orbital period, star temperature, star mass, magnitude and distance.

```

final_df <- compact_df %>% select(pl_name,discoverymethod,OrbitalPeriod,StarMass,Distance,Magnitude,host
final_df <- final_df %>% drop_na(pl_name,discoverymethod,OrbitalPeriod,StarMass,Distance,Magnitude,host
mdf <- missing_data.frame(as.data.frame(final_df))
image(mdf)

```

Dark represents missing data



Clustered by missingness

Converting discovery method into factors

```
final_df <- final_df %>% mutate(discoverymethod = factor(discoverymethod))
summary(final_df)
```

```
##      pl_name                      discoverymethod OrbitalPeriod
##  Length:3351      Transit          :2613   Min.   :    0
##  Class  :character    Radial Velocity   : 709   1st Qu.:     4
##  Mode   :character    Imaging         : 10    Median :    10
##                                         Eclipse Timing Variations   :  7    Mean   : 120920
##                                         Transit Timing Variations   :  5    3rd Qu.:    40
##                                         Orbital Brightness Modulation:  3    Max.   :4020000000
##                                         (Other)           :  4
##      StarMass          Distance       Magnitude      hostname
##  Min.   : 0.0900  Min.   : 1.301  Min.   : 2.364  Length:3351
##  1st Qu.: 0.8100  1st Qu.:111.590  1st Qu.:10.339  Class  :character
##  Median  : 0.9700  Median : 438.359  Median :13.390  Mode   :character
##  Mean    : 0.9818  Mean   :558.506  Mean   :12.377
##  3rd Qu.: 1.1000  3rd Qu.:860.380  3rd Qu.:14.866
##  Max.   :10.9400  Max.   :4483.050  Max.   :19.879
##
```

```
tally(~discoverymethod, data = final_df)
```

```
## discoverymethod
```

```

##          Astrometry      Eclipse Timing Variations
##                2                      7
##          Imaging Orbital Brightness Modulation
##                10                     3
## Pulsation Timing Variations           Radial Velocity
##                2                      709
##          Transit     Transit Timing Variations
##                2613                     5

final_df <- final_df %>% mutate(discoverymethod = forcats::fct_collapse(discoverymethod, "Timing Variations"))
tally(~discoverymethod, data = final_df)

## discoverymethod
##          Other Timing Variations   Radial Velocity           Transit
##                15                  14                 709             2613

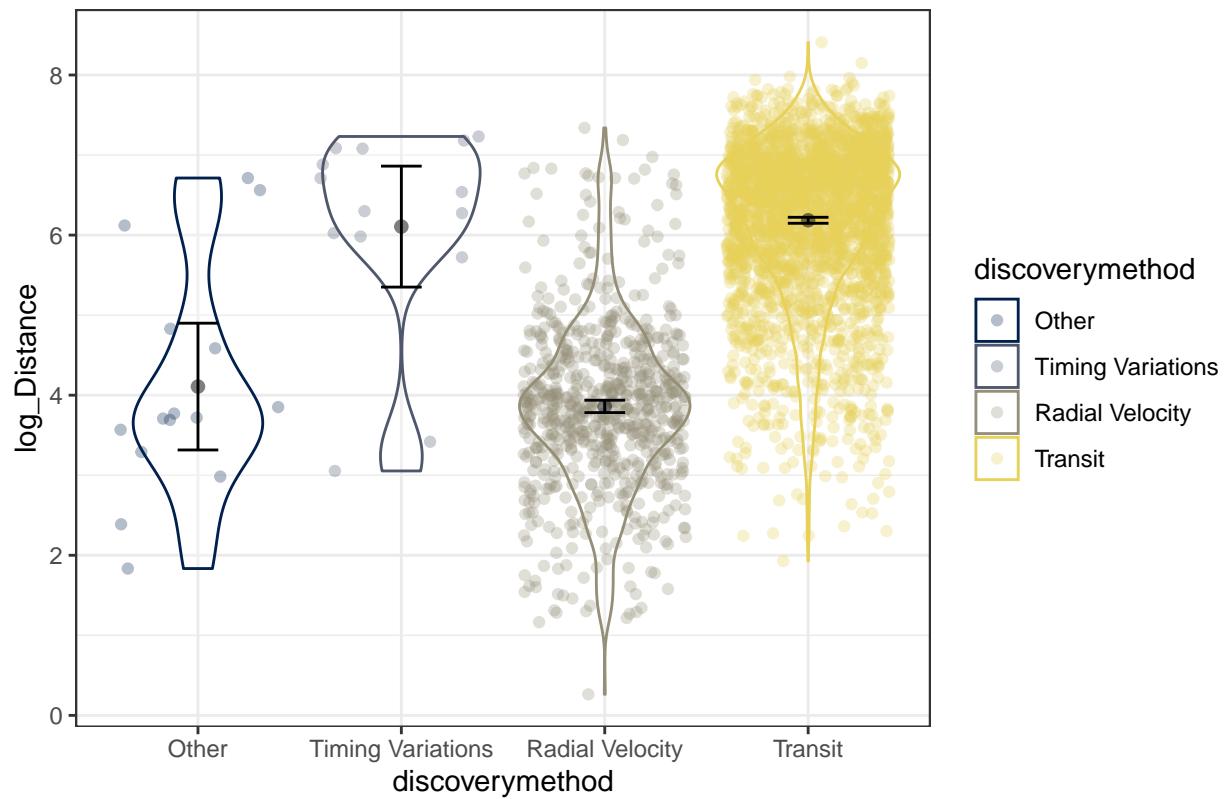
summary(final_df)

##    pl_name            discoverymethod  OrbitalPeriod
##  Length:3351        Other           : 15   Min.   :     0
##  Class :character  Timing Variations: 14   1st Qu.:     4
##  Mode  :character  Radial Velocity : 709  Median :    10
##                    Transit          :2613  Mean   : 120920
##                                         3rd Qu.:    40
##                                         Max.   :402000000
##    StarMass       Distance      Magnitude      hostname
##  Min.   : 0.0900  Min.   : 1.301  Min.   : 2.364  Length:3351
##  1st Qu.: 0.8100  1st Qu.:111.590  1st Qu.:10.339  Class  :character
##  Median : 0.9700  Median : 438.359  Median :13.390  Mode   :character
##  Mean   : 0.9818  Mean   : 558.506  Mean   :12.377
##  3rd Qu.: 1.1000  3rd Qu.: 860.380  3rd Qu.:14.866
##  Max.   :10.9400  Max.   :4483.050  Max.   :19.879

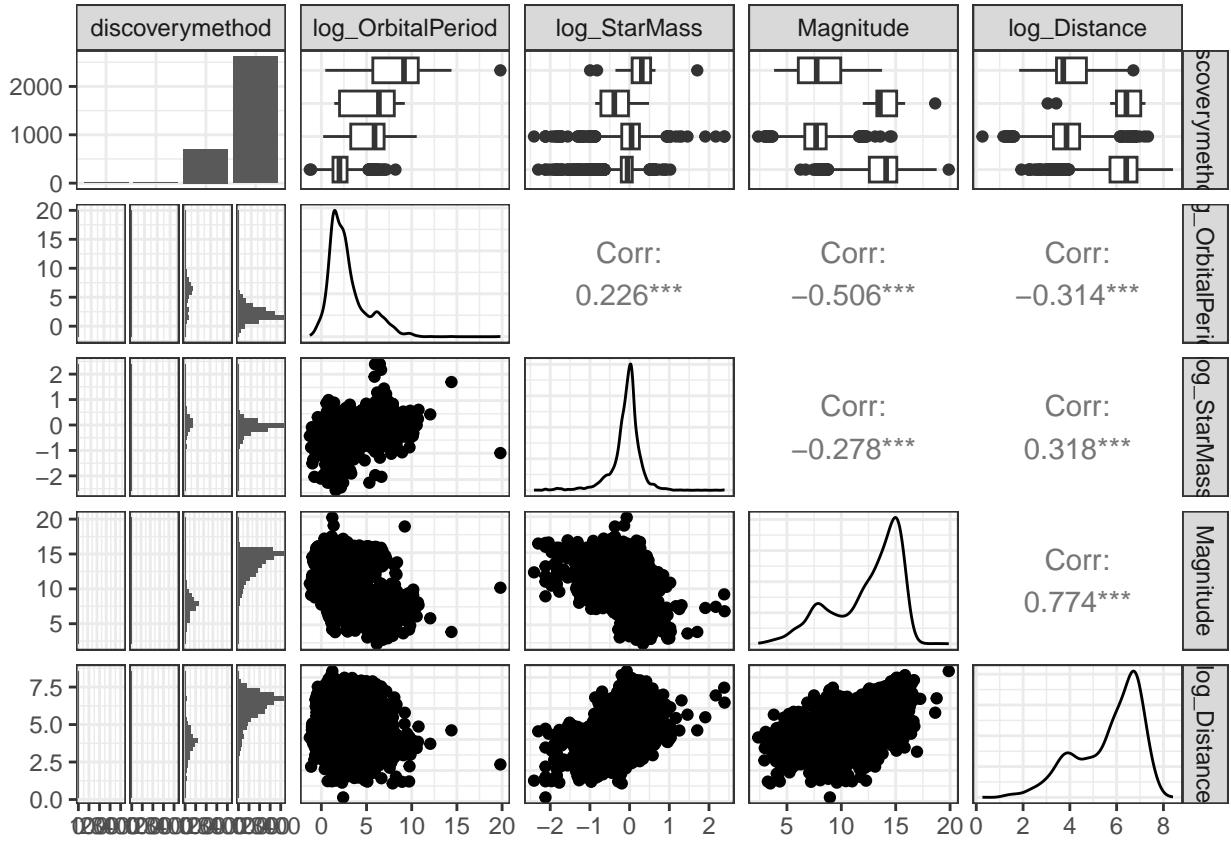
final_df <- final_df %>%
  mutate(log_Distance = log(Distance), log_OrbitalPeriod = log(OrbitalPeriod), log_StarMass = log(StarMass))
enhanced_stripchart(data=final_df,log_Distance~discoverymethod)

```

Enhanced stripchart of log_Distance by discoverymethod

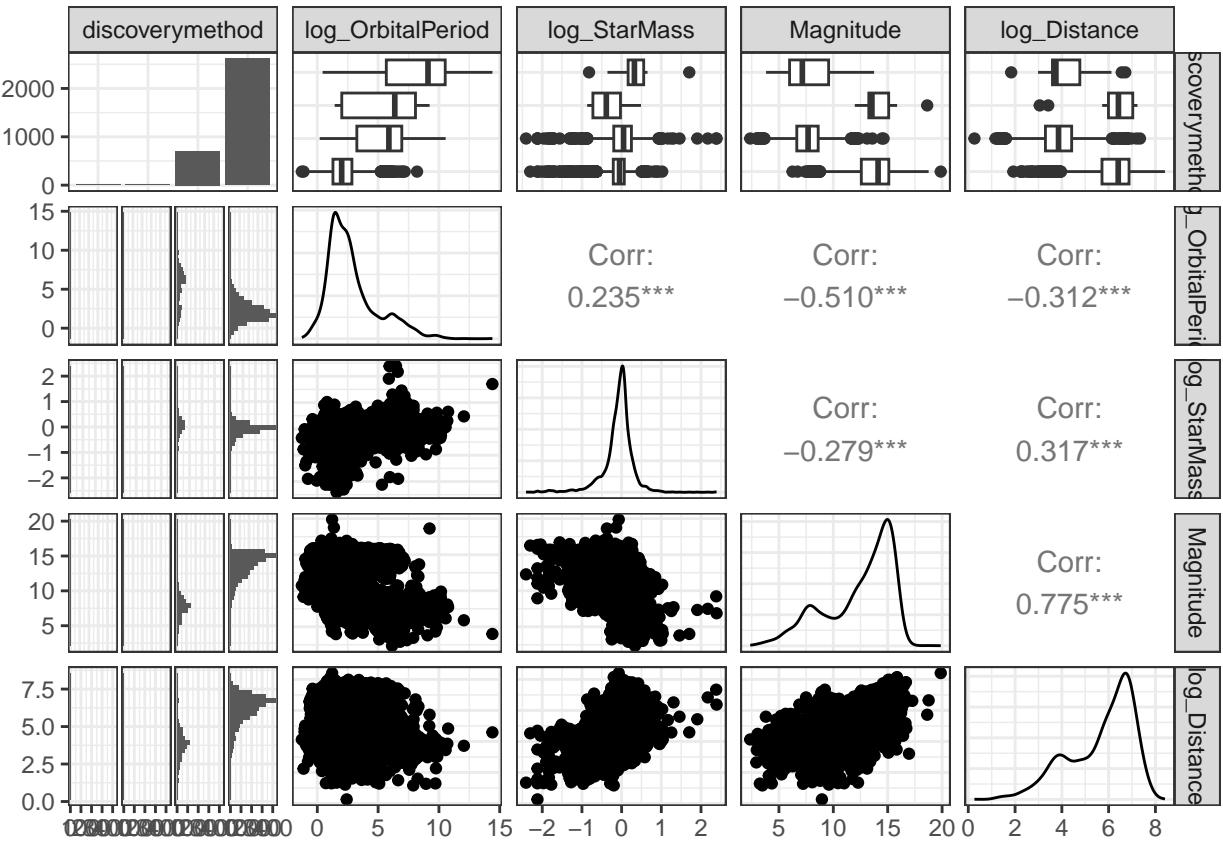


```
library(GGally)
final_df %>% select(discoverymethod,log_OrbitalPeriod,log_StarMass,Magnitude,log_Distance) %>% ggpairs()
```



Remove the outlier in the orbital period since it has very large error

```
final_df <- final_df %>% filter(OrbitalPeriod != 402000000)
final_df %>% select(discoverymethod,log_OrbitalPeriod,log_StarMass,Magnitude,log_Distance) %>% ggpairs()
```



```
summary(final_df)
```

```
##      pl_name          discoverymethod OrbitalPeriod
##  Length:3350      Other           : 14   Min.   : 0.3
##  Class :character Timing Variations: 14   1st Qu.: 4.1
##  Mode  :character Radial Velocity  :709   Median  :10.1
##                  Transit        :2613   Mean    :956.1
##                                         3rd Qu.:39.4
##                                         Max.  :1790000.0
##      StarMass       Distance        Magnitude      hostname
##  Min.   : 0.090   Min.   : 1.301   Min.   : 2.364   Length:3350
##  1st Qu.: 0.810   1st Qu.:111.643   1st Qu.:10.352   Class  :character
##  Median : 0.970   Median : 438.385   Median :13.391   Mode   :character
##  Mean   : 0.982   Mean   : 558.670   Mean   :12.378
##  3rd Qu.: 1.100   3rd Qu.:860.386   3rd Qu.:14.866
##  Max.   :10.940   Max.   :4483.050   Max.   :19.879
##      log_Distance  log_OrbitalPeriod log_StarMass
##  Min.   :0.2633   Min.   :-1.272    Min.   :-2.40795
##  1st Qu.:4.7153   1st Qu.: 1.406    1st Qu.:-0.21072
##  Median :6.0831   Median : 2.316    Median :-0.03046
##  Mean   :5.6845   Mean   : 2.841    Mean   :-0.08549
##  3rd Qu.:6.7574   3rd Qu.: 3.675    3rd Qu.: 0.09531
##  Max.   :8.4081   Max.   :14.398    Max.   : 2.39243
```

Initial model

```

library(lme4)
initial_model <- lm(log_Distance ~ (discoverymethod + log_OrbitalPeriod + log_StarMass + Magnitude), data = final_df)
summary(initial_model)

##
## Call:
## lm(formula = log_Distance ~ (discoverymethod + log_OrbitalPeriod +
##     log_StarMass + Magnitude), data = final_df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -3.1485 -0.2130 -0.0461  0.1597  3.1902 
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)             -0.037353  0.120267 -0.311   0.756    
## discoverymethodTiming Variations  0.736991  0.158277  4.656 3.34e-06  
## discoverymethodRadial Velocity  0.544247  0.112039  4.858 1.24e-06  
## discoverymethodTransit    0.648664  0.115816  5.601 2.31e-08  
## log_OrbitalPeriod        0.035486  0.004638  7.652 2.58e-14  
## log_StarMass              2.072533  0.020600 100.607 < 2e-16  
## Magnitude                0.418006  0.004193  99.685 < 2e-16  
##
## Residual standard error: 0.4127 on 3343 degrees of freedom
## Multiple R-squared:  0.9104, Adjusted R-squared:  0.9102 
## F-statistic:  5662 on 6 and 3343 DF,  p-value: < 2.2e-16

```

```
Anova(initial_model)
```

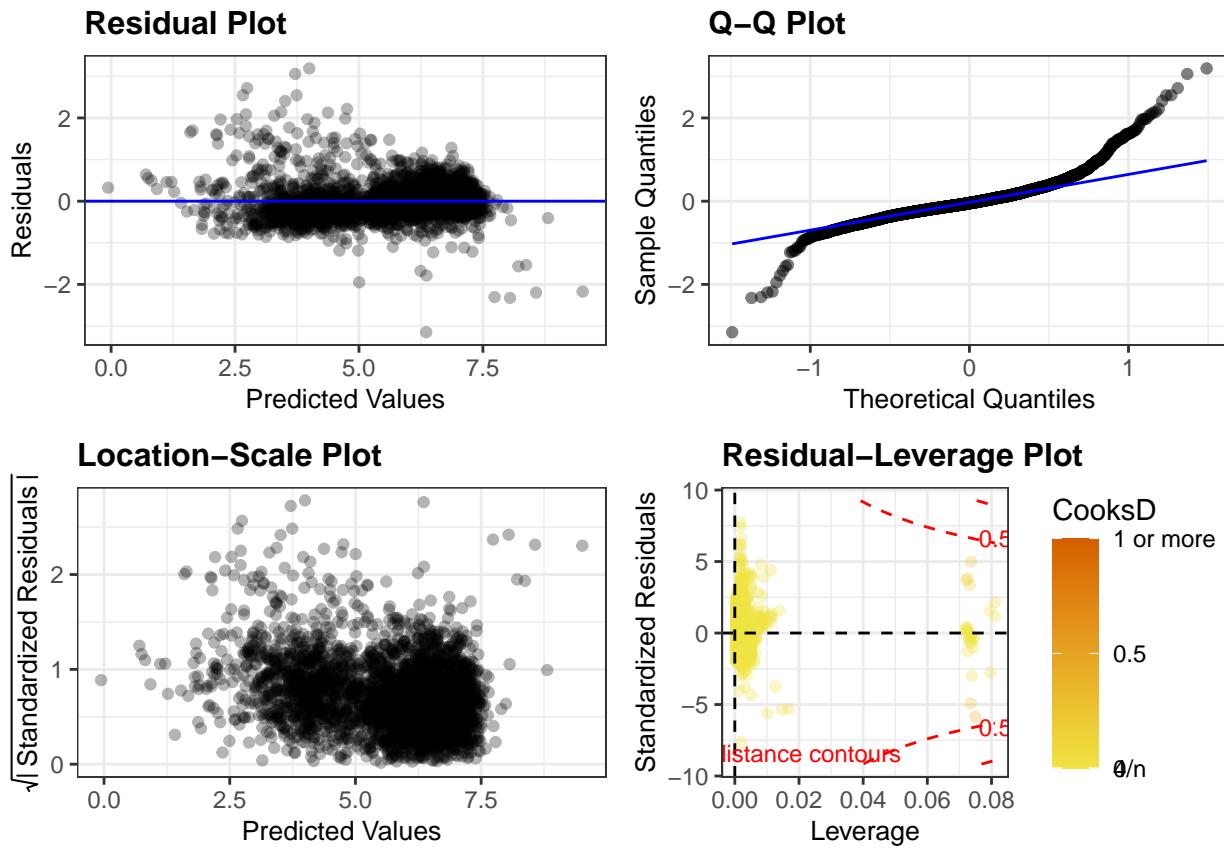
```

## Anova Table (Type II tests)
##
## Response: log_Distance
##                         Sum Sq Df  F value Pr(>F)    
## discoverymethod      6.02   3  11.779 1.128e-07  
## log_OrbitalPeriod    9.97   1   58.547 2.580e-14  
## log_StarMass        1724.05   1 10121.732 < 2.2e-16  
## Magnitude           1692.60   1  9937.100 < 2.2e-16  
## Residuals            569.42 3343

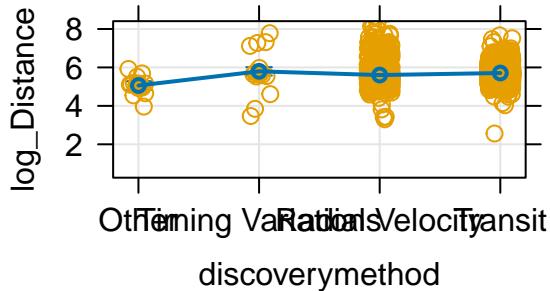
```

Diagnostic and effects plots

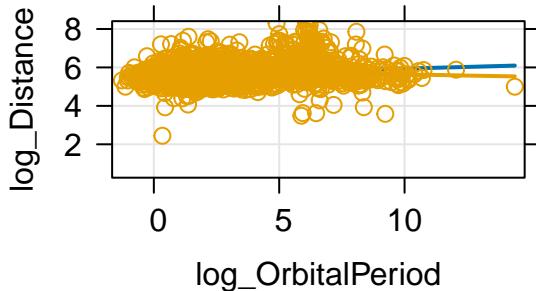
```
resid_panel(initial_model, "R", alpha = 0.3)
```



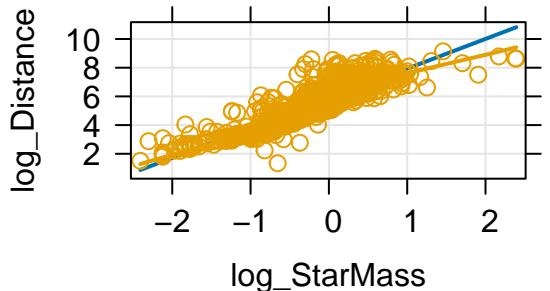
discoverymethod effect plot



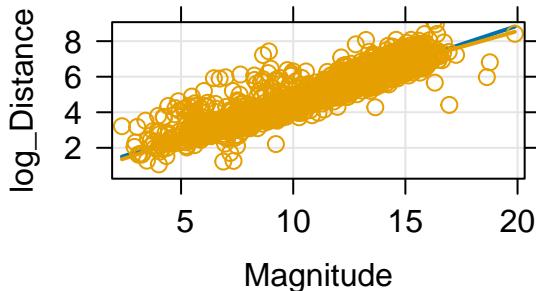
log_OrbitalPeriod effect plot



log_StarMass effect plot



Magnitude effect plot



VIF Calculation

```
vif(initial_model)
```

```
##          GVIF Df GVIF^(1/(2*Df))
## discoverymethod 4.037464 3     1.261880
## log_OrbitalPeriod 1.811232 1     1.345820
## log_StarMass      1.199062 1     1.095017
## Magnitude        3.270176 1     1.808363
```

Four-way interaction model

```
model1 <- lm(log_Distance ~ (discoverymethod*log_OrbitalPeriod*log_StarMass*Magnitude), data=final_df)
summary(model1)
```

```
##
## Call:
## lm(formula = log_Distance ~ (discoverymethod * log_OrbitalPeriod *
##       log_StarMass * Magnitude), data = final_df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.2381 -0.2089 -0.0453  0.1532  3.3213
##
## Coefficients:
```

	Estimate
##	8.450e-01
## (Intercept)	8.450e-01
## discoverymethodTiming Variations	-9.366e+00
## discoverymethodRadial Velocity	7.769e-04
## discoverymethodTransit	-5.291e-01
## log_OrbitalPeriod	5.147e-02
## log_StarMass	1.136e+00
## Magnitude	4.032e-01
## discoverymethodTiming Variations:log_OrbitalPeriod	3.767e+00
## discoverymethodRadial Velocity:log_OrbitalPeriod	7.531e-02
## discoverymethodTransit:log_OrbitalPeriod	1.603e-02
## discoverymethodTiming Variations:log_StarMass	-6.078e+01
## discoverymethodRadial Velocity:log_StarMass	1.426e+00
## discoverymethodTransit:log_StarMass	1.163e+00
## log_OrbitalPeriod:log_StarMass	-7.250e-02
## discoverymethodTiming Variations:Magnitude	6.641e-01
## discoverymethodRadial Velocity:Magnitude	-2.943e-02
## discoverymethodTransit:Magnitude	3.482e-02
## log_OrbitalPeriod:Magnitude	-1.505e-02
## log_StarMass:Magnitude	1.183e-01
## discoverymethodTiming Variations:log_OrbitalPeriod:log_StarMass	1.262e+01
## discoverymethodRadial Velocity:log_OrbitalPeriod:log_StarMass	-3.571e-02
## discoverymethodTransit:log_OrbitalPeriod:log_StarMass	1.171e-01
## discoverymethodTiming Variations:log_OrbitalPeriod:Magnitude	-2.604e-01
## discoverymethodRadial Velocity:log_OrbitalPeriod:Magnitude	2.729e-03
## discoverymethodTransit:log_OrbitalPeriod:Magnitude	1.354e-02
## discoverymethodTiming Variations:log_StarMass:Magnitude	3.941e+00
## discoverymethodRadial Velocity:log_StarMass:Magnitude	-1.835e-01
## discoverymethodTransit:log_StarMass:Magnitude	-1.364e-01
## log_OrbitalPeriod:log_StarMass:Magnitude	1.309e-02
## discoverymethodTiming Variations:log_OrbitalPeriod:log_StarMass:Magnitude	-8.829e-01
## discoverymethodRadial Velocity:log_OrbitalPeriod:log_StarMass:Magnitude	-2.657e-03
## discoverymethodTransit:log_OrbitalPeriod:log_StarMass:Magnitude	-1.328e-02
##	Std. Error
## (Intercept)	3.756e+00
## discoverymethodTiming Variations	6.061e+00
## discoverymethodRadial Velocity	3.761e+00
## discoverymethodTransit	3.758e+00
## log_OrbitalPeriod	3.342e-01
## log_StarMass	5.394e+00
## Magnitude	2.746e-01
## discoverymethodTiming Variations:log_OrbitalPeriod	9.778e-01
## discoverymethodRadial Velocity:log_OrbitalPeriod	3.362e-01
## discoverymethodTransit:log_OrbitalPeriod	3.386e-01
## discoverymethodTiming Variations:log_StarMass	1.615e+01
## discoverymethodRadial Velocity:log_StarMass	5.408e+00
## discoverymethodTransit:log_StarMass	5.403e+00
## log_OrbitalPeriod:log_StarMass	2.575e-01
## discoverymethodTiming Variations:Magnitude	4.261e-01
## discoverymethodRadial Velocity:Magnitude	2.757e-01
## discoverymethodTransit:Magnitude	2.748e-01
## log_OrbitalPeriod:Magnitude	2.576e-02
## log_StarMass:Magnitude	4.246e-01
## discoverymethodTiming Variations:log_OrbitalPeriod:log_StarMass	2.695e+00

## discoverymethodRadial Velocity:log_OrbitalPeriod:log_StarMass	2.665e-01
## discoverymethodTransit:log_OrbitalPeriod:log_StarMass	3.107e-01
## discoverymethodTiming Variations:log_OrbitalPeriod:Magnitude	6.969e-02
## discoverymethodRadial Velocity:log_OrbitalPeriod:Magnitude	2.617e-02
## discoverymethodTransit:log_OrbitalPeriod:Magnitude	2.606e-02
## discoverymethodTiming Variations:log_StarMass:Magnitude	1.121e+00
## discoverymethodRadial Velocity:log_StarMass:Magnitude	4.263e-01
## discoverymethodTransit:log_StarMass:Magnitude	4.252e-01
## log_OrbitalPeriod:log_StarMass:Magnitude	3.254e-02
## discoverymethodTiming Variations:log_OrbitalPeriod:log_StarMass:Magnitude	1.913e-01
## discoverymethodRadial Velocity:log_OrbitalPeriod:log_StarMass:Magnitude	3.330e-02
## discoverymethodTransit:log_OrbitalPeriod:log_StarMass:Magnitude	3.488e-02
##	t value
## (Intercept)	0.225
## discoverymethodTiming Variations	-1.545
## discoverymethodRadial Velocity	0.000
## discoverymethodTransit	-0.141
## log_OrbitalPeriod	0.154
## log_StarMass	0.211
## Magnitude	1.468
## discoverymethodTiming Variations:log_OrbitalPeriod	3.852
## discoverymethodRadial Velocity:log_OrbitalPeriod	0.224
## discoverymethodTransit:log_OrbitalPeriod	0.047
## discoverymethodTiming Variations:log_StarMass	-3.765
## discoverymethodRadial Velocity:log_StarMass	0.264
## discoverymethodTransit:log_StarMass	0.215
## log_OrbitalPeriod:log_StarMass	-0.282
## discoverymethodTiming Variations:Magnitude	1.559
## discoverymethodRadial Velocity:Magnitude	-0.107
## discoverymethodTransit:Magnitude	0.127
## log_OrbitalPeriod:Magnitude	-0.584
## log_StarMass:Magnitude	0.279
## discoverymethodTiming Variations:log_OrbitalPeriod:log_StarMass	4.682
## discoverymethodRadial Velocity:log_OrbitalPeriod:log_StarMass	-0.134
## discoverymethodTransit:log_OrbitalPeriod:log_StarMass	0.377
## discoverymethodTiming Variations:log_OrbitalPeriod:Magnitude	-3.736
## discoverymethodRadial Velocity:log_OrbitalPeriod:Magnitude	0.104
## discoverymethodTransit:log_OrbitalPeriod:Magnitude	0.520
## discoverymethodTiming Variations:log_StarMass:Magnitude	3.514
## discoverymethodRadial Velocity:log_StarMass:Magnitude	-0.430
## discoverymethodTransit:log_StarMass:Magnitude	-0.321
## log_OrbitalPeriod:log_StarMass:Magnitude	0.402
## discoverymethodTiming Variations:log_OrbitalPeriod:log_StarMass:Magnitude	-4.615
## discoverymethodRadial Velocity:log_OrbitalPeriod:log_StarMass:Magnitude	-0.080
## discoverymethodTransit:log_OrbitalPeriod:log_StarMass:Magnitude	-0.381
##	Pr(> t)
## (Intercept)	0.822019
## discoverymethodTiming Variations	0.122356
## discoverymethodRadial Velocity	0.999835
## discoverymethodTransit	0.888034
## log_OrbitalPeriod	0.877618
## log_StarMass	0.833267
## Magnitude	0.142093
## discoverymethodTiming Variations:log_OrbitalPeriod	0.000119

```

## discoverymethodRadial Velocity:log_OrbitalPeriod          0.822770
## discoverymethodTransit:log_OrbitalPeriod                 0.962252
## discoverymethodTiming Variations:log_StarMass            0.000170
## discoverymethodRadial Velocity:log_StarMass              0.792058
## discoverymethodTransit:log_StarMass                      0.829588
## log_OrbitalPeriod:log_StarMass                          0.778331
## discoverymethodTiming Variations:Magnitude             0.119167
## discoverymethodRadial Velocity:Magnitude                0.914992
## discoverymethodTransit:Magnitude                       0.899150
## log_OrbitalPeriod:Magnitude                           0.559087
## log_StarMass:Magnitude                                0.780484
## discoverymethodTiming Variations:log_OrbitalPeriod:log_StarMass 2.95e-06
## discoverymethodRadial Velocity:log_OrbitalPeriod:log_StarMass 0.893382
## discoverymethodTransit:log_OrbitalPeriod:log_StarMass    0.706288
## discoverymethodTiming Variations:log_OrbitalPeriod:Magnitude 0.000190
## discoverymethodRadial Velocity:log_OrbitalPeriod:Magnitude 0.916945
## discoverymethodTransit:log_OrbitalPeriod:Magnitude      0.603373
## discoverymethodTiming Variations:log_StarMass:Magnitude 0.000447
## discoverymethodRadial Velocity:log_StarMass:Magnitude   0.666861
## discoverymethodTransit:log_StarMass:Magnitude           0.748472
## log_OrbitalPeriod:log_StarMass:Magnitude               0.687559
## discoverymethodTiming Variations:log_OrbitalPeriod:log_StarMass:Magnitude 4.07e-06
## discoverymethodRadial Velocity:log_OrbitalPeriod:log_StarMass:Magnitude 0.936410
## discoverymethodTransit:log_OrbitalPeriod:log_StarMass:Magnitude 0.703460
##
## Residual standard error: 0.4016 on 3318 degrees of freedom
## Multiple R-squared:  0.9158, Adjusted R-squared:  0.915
## F-statistic:  1164 on 31 and 3318 DF,  p-value: < 2.2e-16

```

```
Anova(model1,test.statistic = "F")
```

	Sum Sq	Df
## Anova Table (Type II tests)		
##		
## Response: log_Distance		
##		
## discoverymethod	5.75	3
## log_OrbitalPeriod	8.52	1
## log_StarMass	1559.32	1
## Magnitude	1616.01	1
## discoverymethod:log_OrbitalPeriod	3.01	3
## discoverymethod:log_StarMass	4.74	3
## log_OrbitalPeriod:log_StarMass	0.08	1
## discoverymethod:Magnitude	10.63	3
## log_OrbitalPeriod:Magnitude	0.74	1
## log_StarMass:Magnitude	0.49	1
## discoverymethod:log_OrbitalPeriod:log_StarMass	0.16	3
## discoverymethod:log_OrbitalPeriod:Magnitude	0.63	3
## discoverymethod:log_StarMass:Magnitude	0.21	3
## log_OrbitalPeriod:log_StarMass:Magnitude	0.23	1
## discoverymethod:log_OrbitalPeriod:log_StarMass:Magnitude	3.59	3
## Residuals	535.13	3318
##	F value	Pr(>F)
## discoverymethod	11.8831	9.718e-08
## log_OrbitalPeriod	52.8438	4.483e-13

```

## log_StarMass          9668.2971 < 2.2e-16
## Magnitude            10019.7709 < 2.2e-16
## discoverymethod:log_OrbitalPeriod      6.2207 0.0003287
## discoverymethod:log_StarMass           9.7977 1.967e-06
## log_OrbitalPeriod:log_StarMass         0.5211 0.4704223
## discoverymethod:Magnitude            21.9626 4.430e-14
## log_OrbitalPeriod:Magnitude          4.5625 0.0327519
## log_StarMass:Magnitude              3.0182 0.0824289
## discoverymethod:log_OrbitalPeriod:log_StarMass 0.3261 0.8064929
## discoverymethod:log_OrbitalPeriod:Magnitude 1.3104 0.2691691
## discoverymethod:log_StarMass:Magnitude 0.4372 0.7263813
## log_OrbitalPeriod:log_StarMass:Magnitude 1.3983 0.2370933
## discoverymethod:log_OrbitalPeriod:log_StarMass:Magnitude 7.4111 6.034e-05
## Residuals

```

Backward Step-wise Testing Process

```

model2 <- lm(log_Distance ~ (discoverymethod*log_OrbitalPeriod*log_StarMass*Magnitude) - discoverymethod:log_StarMass)
Anova(model2)

```

	Sum Sq	Df	F value	Pr(>F)
##				
## discoverymethod	5.75	3	11.8147	1.073e-07
## log_OrbitalPeriod	8.52	1	52.5395	5.220e-13
## log_StarMass	1559.32	1	9612.6261	< 2.2e-16
## Magnitude	1616.01	1	9962.0761	< 2.2e-16
## discoverymethod:log_OrbitalPeriod	3.01	3	6.1849	0.0003459
## discoverymethod:log_StarMass	4.74	3	9.7412	2.133e-06
## log_OrbitalPeriod:log_StarMass	0.08	1	0.5181	0.4717028
## discoverymethod:Magnitude	10.63	3	21.8362	5.320e-14
## log_OrbitalPeriod:Magnitude	0.74	1	4.5363	0.0332575
## log_StarMass:Magnitude	0.49	1	3.0008	0.0833164
## discoverymethod:log_OrbitalPeriod:log_StarMass	0.16	3	0.3242	0.8078556
## discoverymethod:log_OrbitalPeriod:Magnitude	0.63	3	1.3029	0.2716865
## discoverymethod:log_StarMass:Magnitude	0.21	3	0.4347	0.7281726
## log_OrbitalPeriod:log_StarMass:Magnitude	0.23	1	1.3902	0.2384477
## Residuals	538.72	3321		

```

model3 <- lm(log_Distance ~ (discoverymethod*log_OrbitalPeriod*log_StarMass*Magnitude) - discoverymethod:log_StarMass)
Anova(model3)

```

	Sum Sq	Df	F value	Pr(>F)
##				
## discoverymethod	5.75	3	11.8207	1.063e-07
## log_OrbitalPeriod	8.55	1	52.7594	4.674e-13
## log_StarMass	1559.32	1	9617.5330	< 2.2e-16
## Magnitude	1616.01	1	9967.1613	< 2.2e-16
## discoverymethod:log_OrbitalPeriod	2.76	3	5.6772	0.0007099
## discoverymethod:log_StarMass	4.74	3	9.7462	2.118e-06

```

## log_OrbitalPeriod:log_StarMass           0.02      1   0.1308  0.7175834
## discoverymethod:Magnitude              10.63      3  21.8473 5.233e-14
## log_OrbitalPeriod:Magnitude            0.72      1   4.4274  0.0354414
## log_StarMass:Magnitude                 0.49      1   3.0023  0.0832377
## discoverymethod:log_OrbitalPeriod:log_StarMass  0.70      3  1.4461  0.2273947
## discoverymethod:log_OrbitalPeriod:Magnitude    0.67      3  1.3697  0.2501062
## log_OrbitalPeriod:log_StarMass:Magnitude     0.14      1   0.8775  0.3489420
## Residuals                           538.93  3324

```

```

model4 <- lm(log_Distance~(discoverymethod*log_OrbitalPeriod*log_StarMass*Magnitude)-discoverymethod:log_
Anova(model4)

```

```

## Anova Table (Type II tests)
##
## Response: log_Distance
##                                     Sum Sq Df F value Pr(>F)
## discoverymethod                  5.75  3 11.8160 1.070e-07
## log_OrbitalPeriod                8.55  1 52.7382 4.723e-13
## log_StarMass                     1559.32 1 9613.6662 < 2.2e-16
## Magnitude                        1618.48 1 9978.3608 < 2.2e-16
## discoverymethod:log_OrbitalPeriod 2.76  3  5.6749 0.0007122
## discoverymethod:log_StarMass      4.74  3  9.7423 2.130e-06
## log_OrbitalPeriod:log_StarMass   0.02  1   0.1308  0.7176377
## discoverymethod:Magnitude       10.68  3  21.9419 4.561e-14
## log_OrbitalPeriod:Magnitude     0.61  1   3.7881  0.0517013
## log_StarMass:Magnitude          0.26  1   1.6047  0.2053214
## discoverymethod:log_OrbitalPeriod:Magnitude  0.84  3  1.7335  0.1579033
## log_OrbitalPeriod:log_StarMass:Magnitude  1.08  1  6.6719  0.0098370
## Residuals                         539.63 3327

```

```

model5 <- lm(log_Distance~(discoverymethod*log_OrbitalPeriod*log_StarMass*Magnitude)-discoverymethod:log_
Anova(model5)

```

```

## Anova Table (Type II tests)
##
## Response: log_Distance
##                                     Sum Sq Df F value Pr(>F)
## discoverymethod                  5.75  3 11.8082 1.083e-07
## log_OrbitalPeriod                8.55  1 52.7034 4.806e-13
## log_StarMass                     1565.62 1 9646.0949 < 2.2e-16
## Magnitude                        1618.48 1 9971.7709 < 2.2e-16
## discoverymethod:log_OrbitalPeriod 2.76  3  5.6712 0.0007159
## discoverymethod:log_StarMass      5.37  3  11.0304 3.327e-07
## log_OrbitalPeriod:log_StarMass   0.06  1   0.3825  0.5362854
## discoverymethod:Magnitude       10.68  3  21.9274 4.657e-14
## log_OrbitalPeriod:Magnitude     0.61  1   3.7856  0.0517785
## log_StarMass:Magnitude          0.32  1   1.9519  0.1624799
## log_OrbitalPeriod:log_StarMass:Magnitude  0.98  1  6.0351  0.0140745
## Residuals                         540.48 3330

```

```

summary(model5)

```

```

## 
## Call:
## lm(formula = log_Distance ~ (discoverymethod * log_OrbitalPeriod *
##     log_StarMass * Magnitude) - discoverymethod:log_OrbitalPeriod:log_StarMass:Magnitude -
##     discoverymethod:log_StarMass:Magnitude - discoverymethod:log_OrbitalPeriod:log_StarMass -
##     discoverymethod:log_OrbitalPeriod:Magnitude, data = final_df)
##
## Residuals:
##      Min    1Q Median    3Q   Max 
## -3.2890 -0.2107 -0.0459  0.1571  3.3088 
##
## Coefficients:
##                               Estimate Std. Error t value
## (Intercept)                0.775387  1.124323  0.690
## discoverymethodTiming Variations        4.061341  1.473745  2.756
## discoverymethodRadial Velocity        0.327390  1.111157  0.295
## discoverymethodTransit          -0.558083  1.113432 -0.501
## log_OrbitalPeriod           -0.025464  0.071083 -0.358
## log_StarMass                  2.539905  0.348219  7.294
## Magnitude                   0.411912  0.083493  4.933
## discoverymethodTiming Variations:log_OrbitalPeriod  0.038972  0.077911  0.500
## discoverymethodRadial Velocity:log_OrbitalPeriod   0.099426  0.064581  1.540
## discoverymethodTransit:log_OrbitalPeriod          0.148913  0.064101  2.323
## discoverymethodTiming Variations:log_StarMass      -1.017210  0.417950 -2.434
## discoverymethodRadial Velocity:log_StarMass        -0.314397  0.272600 -1.153
## discoverymethodTransit:log_StarMass                 0.034366  0.277715  0.124
## log_OrbitalPeriod:log_StarMass                  -0.062040  0.030891 -2.008
## discoverymethodTiming Variations:Magnitude       -0.266749  0.104292 -2.558
## discoverymethodRadial Velocity:Magnitude          -0.070464  0.082199 -0.857
## discoverymethodTransit:Magnitude                 0.033268  0.082593  0.403
## log_OrbitalPeriod:Magnitude                   -0.005542  0.002879 -1.925
## log_StarMass:Magnitude                      -0.037630  0.013729 -2.741
## log_OrbitalPeriod:log_StarMass:Magnitude       0.007201  0.002931  2.457
## Pr(>|t|)                                 0.49046
## (Intercept)                            0.00589
## discoverymethodTiming Variations        0.76829
## discoverymethodRadial Velocity          0.61624
## discoverymethodTransit                  0.72019
## log_OrbitalPeriod                     3.74e-13
## log_StarMass                         8.47e-07
## discoverymethodTiming Variations:log_OrbitalPeriod  0.61696
## discoverymethodRadial Velocity:log_OrbitalPeriod  0.12377
## discoverymethodTransit:log_OrbitalPeriod      0.02023
## discoverymethodTiming Variations:log_StarMass      0.01499
## discoverymethodRadial Velocity:log_StarMass        0.24886
## discoverymethodTransit:log_StarMass          0.90152
## log_OrbitalPeriod:log_StarMass            0.04469
## discoverymethodTiming Variations:Magnitude      0.01058
## discoverymethodRadial Velocity:Magnitude        0.39138
## discoverymethodTransit:Magnitude             0.68712
## log_OrbitalPeriod:Magnitude               0.05432
## log_StarMass:Magnitude                  0.00616
## log_OrbitalPeriod:log_StarMass:Magnitude     0.01407

```

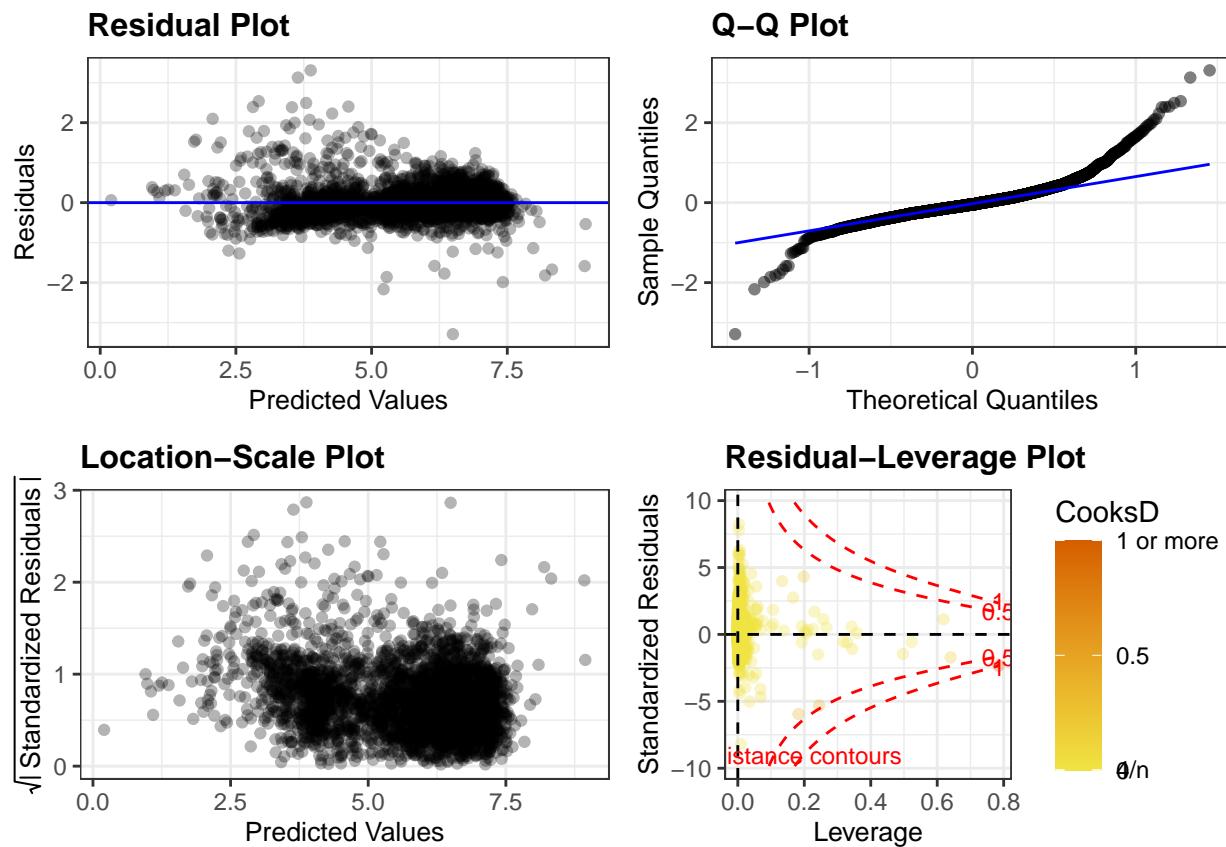
```

## 
## Residual standard error: 0.4029 on 3330 degrees of freedom
## Multiple R-squared:  0.915, Adjusted R-squared:  0.9145
## F-statistic:  1886 on 19 and 3330 DF, p-value: < 2.2e-16

```

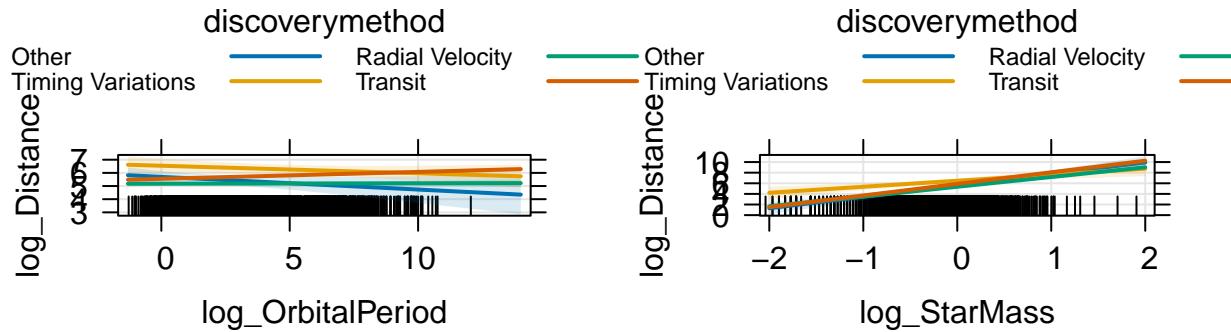
Diagnostics and effects for final model

```
resid_panel(model5, "R", alpha = 0.3)
```

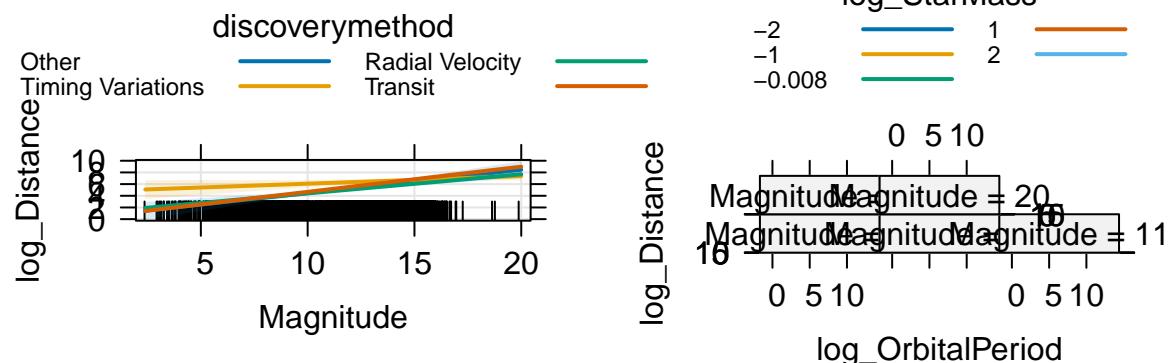


```
plot(allEffects(model5), grid = T, multiline=T, ci.style="bands")
```

verymethod*log_OrbitalPeriod effect plot



discoverymethod*log_StarMass effect plot



```
#plot(allEffects(model5), grid = T, multiline=T, ci.style="bands", selection=1)
#plot(allEffects(model5), grid = T, multiline=T, ci.style="bands", selection=2)
#plot(allEffects(model5), grid = T, multiline=T, ci.style="bands", selection=3)
#plot(allEffects(model5), grid = T, multiline=T, ci.style="bands", selection=4)
```

Pairwise comparison between the discovery methods in original additive model

```
library(emmeans)
res1 <- emmeans(initial_model, pairwise ~ discoverymethod, adjust="tukey")
res1
```

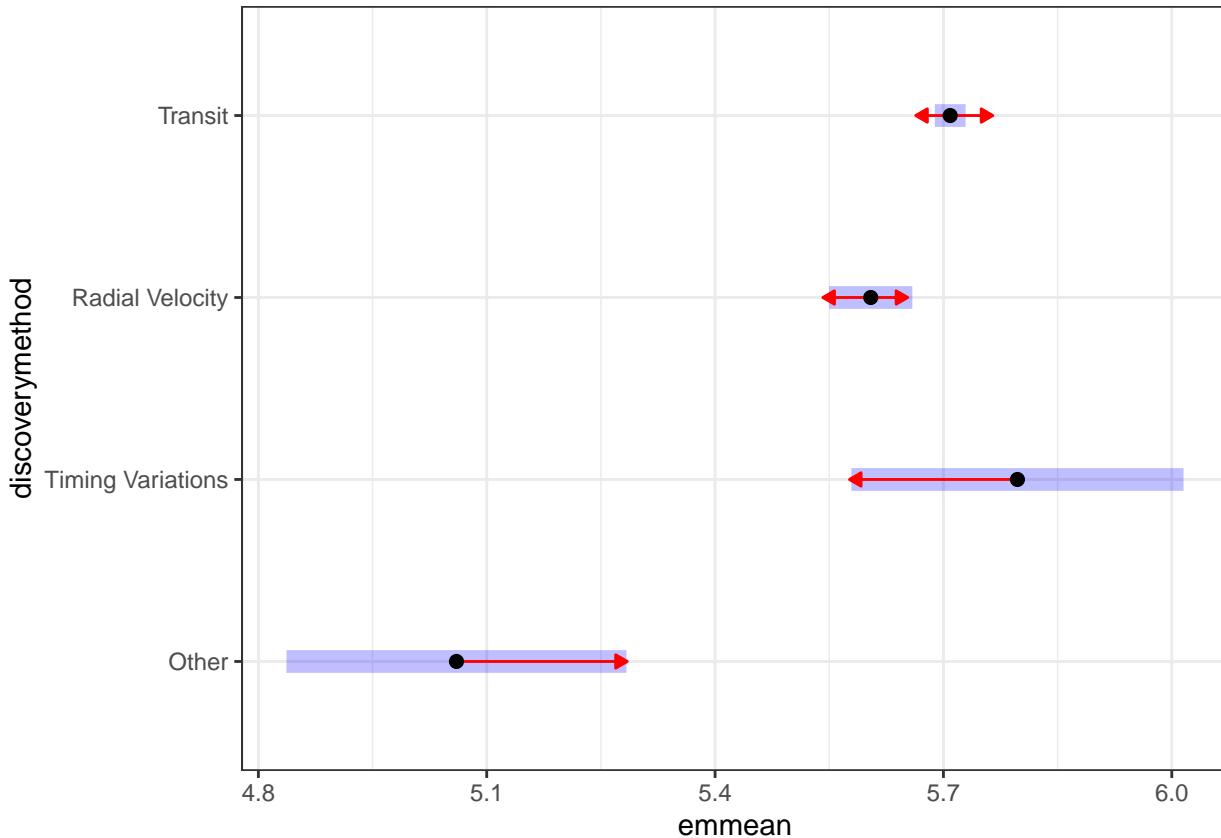
```
## $emmeans
##  discoverymethod    emmean      SE   df lower.CL upper.CL
##  Other             5.060 0.11389 3343    4.837  5.284
##  Timing Variations 5.797 0.11129 3343    5.579  6.015
##  Radial Velocity   5.604 0.02787 3343    5.550  5.659
##  Transit            5.709 0.01032 3343    5.689  5.729
##
##  Confidence level used: 0.95
##
## $contrasts
##  contrast                      estimate      SE   df t.ratio p.value
##  Other - Timing Variations     -0.7370 0.1583 3343  -4.656  <.0001
##  Other - Radial Velocity      -0.5442 0.1120 3343  -4.858  <.0001
##  Other - Transit               -0.6487 0.1158 3343  -5.601  <.0001
```

```

##  Timing Variations - Radial Velocity  0.1927 0.1144 3343  1.685  0.3317
##  Timing Variations - Transit         0.0883 0.1119 3343  0.789  0.8593
##  Radial Velocity - Transit          -0.1044 0.0344 3343 -3.038  0.0128
##
## P value adjustment: tukey method for comparing a family of 4 estimates

```

```
plot(res1, comparison=T)
```



```
multcomp:::cld(res1,alpha=0.05, Letters=LETTERS)
```

```

##  discoverymethod   emmean      SE  df lower.CL upper.CL .group
##  Other            5.060 0.11389 3343    4.837   5.284    A
##  Radial Velocity 5.604 0.02787 3343    5.550   5.659    B
##  Transit          5.709 0.01032 3343    5.689   5.729    C
##  Timing Variations 5.797 0.11129 3343    5.579   6.015   BC
##
## Confidence level used: 0.95
## P value adjustment: tukey method for comparing a family of 4 estimates
## significance level used: alpha = 0.05
## NOTE: If two or more means share the same grouping symbol,
##       then we cannot show them to be different.
##       But we also did not show them to be the same.

```

```

model5.emms <- emmeans(initial_model, "discoverymethod")
model5.emms

##  discoverymethod    emmean      SE   df lower.CL upper.CL
##  Other             5.060 0.11389 3343    4.837   5.284
##  Timing Variations 5.797 0.11129 3343    5.579   6.015
##  Radial Velocity   5.604 0.02787 3343    5.550   5.659
##  Transit            5.709 0.01032 3343    5.689   5.729
##
##  Confidence level used: 0.95

contrast(model5.emms,list(gamma1=c(0,0,-1,1))) %>% confint() %>% as.data.frame()

##  contrast estimate      SE   df lower.CL upper.CL
##  gamma1   0.1044173 0.03436884 3343 0.03703122 0.1718034
##
##  Confidence level used: 0.95

contrast(model5.emms,list(gamma1=c(-1,0,0,1))) %>% confint() %>% as.data.frame()

##  contrast estimate      SE   df lower.CL upper.CL
##  gamma1   0.6486643 0.1158157 3343 0.4215874 0.8757411
##
##  Confidence level used: 0.95

contrast(model5.emms,list(gamma1=c(-1,1,0,0))) %>% confint() %>% as.data.frame()

##  contrast estimate      SE   df lower.CL upper.CL
##  gamma1   0.7369909 0.1582775 3343 0.4266604 1.047321
##
##  Confidence level used: 0.95

contrast(model5.emms,list(gamma1=c(-1,0,0,1))) %>% confint() %>% as.data.frame()

##  contrast estimate      SE   df lower.CL upper.CL
##  gamma1   0.6486643 0.1158157 3343 0.4215874 0.8757411
##
##  Confidence level used: 0.95

```