# SURV 727 Final Project Report

## Ravelry Popularity Exploration

Sarah Glidden

### GitHub link: https://github.com/sarahglidden97/SURV-727-Final-Project

### Introduction

This report investigates what, if any, fiber arts pattern characteristics impact the popularity of a pattern on Ravelry.com. Ravelry is an online pattern database where fiber artists and pattern designers can upload knitting and crochet patterns for sale. The site is also a community space, organizational tool, and yarn database. Many fiber artists make a large portion of their income through pattern sales on Ravelry and their personal sites. Even pattern designers who sell exclusively on their personal sites will still upload pattern pages to Ravelry with a link to their site for payment and downloading. Through the research process, I created a Shiny app tool which could be used by pattern designers to see what characteristics of patterns perform best on Ravelry to increase their popularity and sales, or to find gaps in the pattern market.

### Data

To create a data set for analysis, I used the Hot Right Now Top 20 patterns on the Ravelry Patterns page. Ravelry describes these patterns as "Designs with the most visits in the last 24 hours." I pulled the top 20 Ravelry patterns each day from November 16 to November 25, excluding November 24. I chose to use these patterns to represent the most popular patterns of the day because pattern designers would like to know which patterns are receiving the most traffic.

To obtain this data, I used the Ravelry.com API combined with the ravelRy package in R to pull the top 20 patterns and their characteristics each day. I then created a loop which cleaned each day's data set individually. This included unnesting tables within the data

table and renaming variables. After the data was uniformly cleaned, I bound all data sets to create my data set that would be used for analyses and my Shiny app tool. I then created a `Number_of_Top_20_Appearances` variable to count how often each pattern had appeared over the data collection period.

When the data had been combined, I arranged each pattern by it's most recent pull date and removed all duplicate rows. This was done to ensure that each pattern had the most recent number of likes in it's time in the top 20.

The final data set contained 76 unique patterns. The columns used in the Shiny app tool were:

- `Average_Difficulty_Rating`: a continuous numeric variable, the average user difficulty rating between 0 and 5.
- `Average_Rating`: a numeric variable, the average user rating between 0 and 5.
- `Craft`: a factor variable.
  - Knitting
  - Crochet
- `Free`: a factor variable.
  - Free
  - Paid
- `Knitting_Needle_Size`: a factor variable, the size of knitting needles needed for the pattern.
- `Number_of_Favorites`: a discrete numeric variable, the number of favorites on the pattern on it's most recent day in the Top 20 in the data set.
- `Number_of_Languages`: a discrete numeric variable, the number of languages a pattern is offered in (generated by programmer).
- `Number_of_Top_20_Appearances`: a discrete numeric variable, the number of days the pattern appeared in the Top 20 throughout the data collection period (generated by programmer).
- `Pattern_Category`: a factor variable, the category of craft type (ex. pullover, mittens).
- `Pattern_Parent_Category`: a factor variable, the high-level category of craft type (ex. sweater, hands).
- `Pattern_Grandparent_Category`: a factor variable, the highest-level category of craft type (ex. clothing accessories).

- `Yarn_Weight`: a factor variable, the weight of yarn needed for the pattern.

The columns not used in the Shiny app were:

- `crochet_hook`: a factor variable. This was not used because only one pattern in the final data set was a crochet pattern and it was redundant with the `Craft` variable.
  - TRUE: a crochet hook is needed for the pattern
  - FALSE: a crochet hook is not needed for the pattern
- `crochet_hook_size`: a factor variable, the crochet hook size needed for the pattern. This was not used because only one pattern in the final data set was a crochet pattern.
- `knit_needles`: a factor variable. This was not used because it was redundant with the `Craft` variable.
  - TRUE: knitting needles are needed for the pattern
  - FALSE: knitting needles are not needed for the pattern
- `name`: a categorical variable, the project name. This variable was used to clean the final data set, but was not included in the tool.
- `pull_date`: a date variable, the full date the pattern was pulled from Ravelry. This was used to clean the final data set, but was not included in the tool.
- `pull_day`: a discrete numeric variable, the numeric day the pattern was pulled from Ravelry. This was used to clean the final data set, but was not included in the tool.

## Analysis

To begin my analysis, I planned to use `Number_of_Favorites` and `Number_of_Top_20_Appearances` as my two dependent variables, with all other variables serving as predictors of popularity.

I then ran linear analyses of each of my variables as predictors for the two dependent variables and tested for significance at the $\alpha = 0.05$ level.

### Average_Difficulty_Rating

`Average_Difficulty_Rating` was found to be a statistically significant predictor of `Number_of_Favorites`, with a p-value of 0.000145. The regression indicated that as pattern difficulty increased, the number of favorites on the pattern was expected to increase.

`Average_Difficulty_Rating` was also found to be a statistically significant predictor of `Number_of_Top_20_Appearances`, with a p-value of approximately 0. As the difficulty rating increased, the number of top 20 appearances was expected to increase.

### Average_Rating

`Average_Rating` was found to be a statistically significant predictor of `Number_of_Favorites` with a p-value of approximately 0. As the average rating increased, the number of favorites on the pattern was expected to increase.

`Average_Rating` was also found to be a statistically significant predictor of `Number_of_Top_20_Appearances`, with a p-value of approximately 0. As the average rating increased, the number of top 20 appearances was expected to increase.

### Craft

`Craft` was not found to be a statistically significant predictor of `Number_of_Favorites` or `Number_of_Top_20_Appearances`.

### Free

`Free` was not found to be a statistically significant predictor of `Number_of_Favorites` or `Number_of_Top_20_Appearances`.

### Knitting_Needle_Size

Size 2.5 knitting needles were found to be a statistically significant predictor of `Number_of_Favorites` with a p-value of 0.0447. When a pattern called for size 2.5 needles, the number of favorites was expected to increase.

No knitting needle sizes were statistically significant predictors of `Number_of_Top_20_Appearances`.

### Number_of_Languages

`Number_of_Languages` was found to be a statistically significant predictor of `Number_of_Favorites` with a p-value of approximately 0.000102. As the number of languages a pattern was offered in increased, the number of favorites on the pattern was expected to increase.

`Number_of_Languages` was also found to be a statistically significant predictor of `Number_of_Top_20_Appearances`, with a p-value of approximately 0.000718. As the number of languages a pattern was offered in increased, the number of top 20 appearances was expected to increase.

### Pattern_Category

`Pattern_Category` was not a statistically significant predictor of `Number_of_Favorites`.

Slippers was a statistically significant `Pattern_Category` predictor of `Number_of_Top_20_Appearances` with a p-value of 0.0282. When the pattern category was "slippers," the number of top 20 appearances was expected to increase.

### Pattern_Parent_Category

`Pattern_Parent_Category` was not a statistically significant predictor of `Number_of_Favorites`.

Feet/legs was a statistically significant `Pattern_Parent_Category` predictor of `Number_of_Top_20_Appearance` with a p-value of 0.022. When the pattern category was "feet/legs," the number of top 20 appearances was expected to increase.

### Pattern_Grandparent_Category

`Pattern_Gradparent_Category` was not found to be a statistically significant predictor of `Number_of_Favorites` or `Number_of_Top_20_Appearances`.

### Yarn_Weight

`Yarn_Weight` was not found to be a statistically significant predictor of `Number_of_Favorites`.

All yarn weights were statistically significant predictors of `Number_of_Top_20_Appearances`. Unlisted yarn weight was used as the baseline, and when any other weight of yarn was listed, the number of top 20 appearances was expected to decrease.

### Shiny App

While linear regression provided some impactful insights, not all variables met the assumptions needed for linear analysis. I decided to create a customizable Shiny app which would all pattern designers to customize the x and y axes of bar charts to observe patterns and common themes among the most popular patterns in my analysis window, without drawing statistically inaccurate conclusions. The Shiny app is available to use here: https://sarahglidden97.shinyapps.io/Ravelry_Popularity_Explorer/

## Conclusion

While I would caution against making any statistical conclusions based on the available data, there are some interesting patterns gathered from the Shiny app that are worth noting.

### Average_Difficulty_Rating

Patterns with an average difficulty rating between 2 and 3 had the most favorites, and patterns with unrated difficulty had the most top 20 appearances in the data collection window. These results lead me to believe that while difficulty can impact the number of favorites, it does not impact appearing in the top 20 as harshly.

### Average_Rating

Patterns with an average rating between 4.5 and 5 had the most favorites, and unrated patterns had the most top 20 appearances in the data collection window. These results lead me to believe that while rating can impact the number of favorites, it does not impact appearing in the top 20 as harshly.

### Craft

Knitting projects had the most favorites in the data collection window, while crochet projects had none. Knitting projects appeared on the top 20 list over 150 times, while crochet projects appeared less than 10 times. It appears that people are coming to Ravelry for knitting projects more often than for crochet projects, and that there could be a better space to share crochet projects with more people.

### Free

Paid patterns had more favorites and most top 20 appearances during the data collection window than free patterns. Pattern designers should not be concerned with charging for their patterns impacting their popularity.

### Knitting_Needle_Size

Patterns using knitting needles size 7 had the most favorites during the data collection window, followed by 2.5 and 8. Projects using knitting needles size 7 also had the most appearances in the top 20, followed by 4 and 8.

### Number_of_Languages

Knitting patterns offered in 4 languages had the most favorites during the data collection window, followed by 7 and 10. Knitting patterns offered in 1 language had the most appearances in the top 20, followed by 2 and 4. Generally, offering a pattern in more than one language increases the popularity, but as we can see, a pattern being in one language does not hinder it from being a top 20 pattern.

### Pattern_Category

Pullovers had the most favorites during the data collection window, followed by balaclavas and scarfs. These three categories also had the most appearances in the top 20, in the same order.

### Pattern_Parent_Category

The sweater parent category had the most favorites during the data collection window, followed by neck/torso and hat. These three parent categories also had the most appearances in the top 20, though the order was neck/torso, sweater, then hat.

### Pattern_Grandparent_Category

The accessories grandparent category had the most favorites during the data collection window, followed by clothing and feet/legs. These three grandparent categories also had the most appearances in the top 20, in the same order.

### Yarn_Weight

Patterns using DK yarn had the most favorites during the data collection window, followed by patterns with the yarn weight unlisted and fingering weight yarn. Projects using DK yarn also had the most appearances in the top 20, followed by fingering and worsted weight yarns.

### Limitations

A limitation of this research was the short data collection period. While the patterns in this data may be generalizeable to November and maybe even to Fall and Winter, a larger data collection period including top 20 pattern lists from more months of the year would be needed to make year long conclusions about pattern popularity on Ravelry. It is very likely that sweaters, balaclavas, and hats knit with heavy weight yarn and large knitting needles loom

large this time of year, and that their popularity wanes for other fabrics and styles in the warmer months.

This research also ignores whether pattern creators use social media to promote their Ravelry patterns and how this could impact the likelihood of appearing in Ravelry's top 20. This information appears to be beyond the scope of the Ravelry API and would likely require access to additional data such as pattern designer social media links which may be difficult to access.