Data Visualization: GripTape Research Summary

Christine Shi

Introduction

This document is meant to showcase my skills in data science, data visualization, and the written explanation of data to a non-technical audience. It includes the content pages from an example Research Summary that I made for the GripTape project in the Purpose and Identity Processes Lab at Cornell. The code I used to produce the graphs in the Research Summary is included after the Example Research Summary.

An example Research Summary from my work with the GripTape project in the Purpose and Identity Processes Lab at Cornell.

GripTape is a youth development program that offers high-school-age participants ("Challengers") grants and mentoring to undertake a "Learning Challenge" to learn a project of their choosing. They fill out a survey daily with their scores on a number of key variables, including purpose, perceived Challenge progress, self-esteem, and positive and negative affect. Our lab, in collaboration with GripTape, researches the trajectories Challengers experience throughout their Learning Challenges.

In order to promote open science and a bidirectional flow of information in research, the research assistants on the GripTape project have been creating "Research Summaries," documents with a static cover, introduction, and credit page, along with three individualized content pages sharing Challengers' own data from their Challenge with them with accessible graphs and explanations. These Research Summaries are made from start to finish by the RAs, who take the GripTape data, clean it, explore trends in it, and visualize it in RStudio, write explanations, and send the Summaries off to supervising project leaders for feedback and compilation into the final deliverable. In this supplement, I have only included the content pages with graphs and explanations that I produced.

The Research Summaries have been received by Challengers with enthusiasm and were cited as the second most popular reason for participation in research. Mark Murphy, founder and CEO of GripTape, said, "No one gets data like this. The value added to the young person is gigantic. But the value added to the field as a whole, as I'm talking to funders, other nonprofit partners . . . [Research Summaries] are rattling the assumptions of people leading the field . . . and questioning what they thought was possible."

# Your Psychological Network During Your Learning Challenge

In this Research Summary, we're going to look at how the five variables listed in the following section interacted with each other during your Learning Challenge. Then we're going to dive in and take a closer look at 2-3 things in particular based on what we see.

What Did We Study?

1. **Perceived Challenge progress** describes feeling like you are moving towards an end goal. It is doing your best to be better today than you were yesterday. When making progress, people feel accomplished after working hard at something. Knowing exactly what you want to achieve is important because it lets you measure your progress.

2. **Motivation** is your inner drive to want to do things and finish them, even when those things get hard. Motivation can help inspire you to set goals and follow through with them. Motivated people tend to feel more in control of their lives, become more successful, and report being happier!

3. **Life satisfaction** is your general feeling about your life and how pleased you are with how it's going. It considers looking at your life as a whole rather than your current level of happiness. There are many things that contribute to life satisfaction. Some examples include relationships with family and friends, personal growth, and health & wellness.
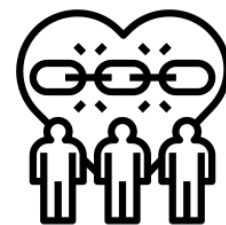

4. **Positive affect.** Affect has to do with your mood and emotions. When you're feeling happy, content, relaxed, or other positive emotions, you have a high positive affect. Feeling these positive emotions can make your days brighter and contribute to your positive attitude. It can also lead to a number of benefits, like a greater sense of purpose and meaning in life, less pain when injured, and better relationships with others!
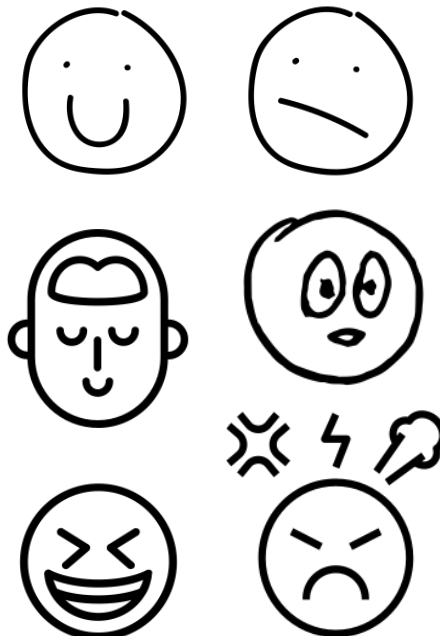
5. **Negative affect.** When you're feeling down, anxious, angry, or any other type of blue, you're experiencing negative affect. While negative affect doesn't sound too fun, it's important to be aware of and experience all your emotions and feelings. These emotions can tell us a lot about ourselves — maybe they are telling us we don't like the situation we are in, and we should find a way to change it up. When we are in tune with these emotions and keeping track of them, they can be super helpful in figuring out what works for us best.

Reading A Network Graph

In social science, we know that different thoughts and feelings influence one another. On a day you're feeling tired, you might be less energized to pursue your goals. On a day you're happy, you might be feeling more creative or excited about pursuing what you love. These are called **associations**. On the next page is a fancy graph, called a **network.** A network is one way we can summarize the associations between variables like your **1)** perceived Challenge progress, **2)** motivation, **3)** life satisfaction, **4)** positive affect, and **5)** negative affect during your Learning Challenge. Everyone's network looks completely different, and this one is specifically yours. We were able to generate this special network for you because of how diligent you were with completing your daily surveys—you gave us so much information! Thank you!
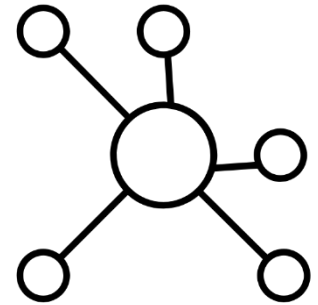
The circles (**nodes**) in the graph represent the variables. The arrows indicate the type, timing, and strength of the associations between the variables. Let's go step-by-step and take a look at what the arrows mean:

Green vs. Red = Positive vs. Negative

Green means positive (higher levels of one variable are associated with higher levels of the connected variable) and red means negative (higher levels of one are associated with lower levels of the connected variable).

Solid vs. Dashed = Same-day vs. Next-day

A *solid* line means an association between the two connected variables on the *same day* (higher levels of one variable predict higher or lower levels of another variable on the *same* day depending on the color of the line).

In contrast, a *dashed* line means an association between one variable (where the arrow starts) and the connected variable (where the arrow is pointing) on the *next day*. In other words, higher levels of one variable predict higher or lower levels of the connected variable on the *next* day, depending on color of the line.
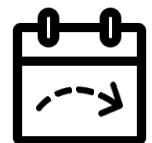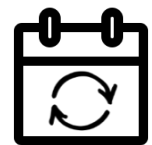
Thickness = Strength of the Association

The wider the arrow, the more closely associated the variables are. It tells us how likely we are to know something about the second variable (where the arrow is pointing) based on what we know about how you scored on the first variable (where the arrow starts).

Loops = Stability

The dashed **loops** around each node represent how well each variable predicts itself the next day. For variables with thick loops, it means we have a pretty good idea of where you'll score on that variable tomorrow as long as we know your score today. For variables with thin loops, it means we are less sure of how you'll score on that variable from day-to-day. Sometimes we call this quantity the **stability** of the variable. Some variables are naturally more stable than others.

Your Psychological Network

Now that we've learned how to read a network graph, let's figure out what yours says about your perceived Challenge progress, motivation, life satisfaction, positive affect, and negative affect throughout your Learning Challenge.

Perceived Challenge progress and motivation: The dashed green loops around each node mean that both variables were quite stable: high perceived Challenge progress tended to predict high perceived Challenge progress the next day, and high motivation on one day tended to predict high motivation for you the next day. There is a solid green line (your thickest line!) between perceived Challenge progress and motivation, which means that on days when you felt like you were making more progress on your Challenge than was usual for you, you were also likely to feel more motivated.

Life satisfaction: The dashed green loop around node 3 shows that higher life satisfaction on one day predicted higher life satisfaction on the next day. The solid green line going to motivation shows that on days when you were more satisfied with your life, you also tended to feel more motivated.

Affect and perceived Challenge progress: Next, notice the dashed red line running from node 5 (negative affect) to node 1 (perceived Challenge progress). This means that when you felt more negative affect on a given day, you tended to feel like you were making less progress on your Challenge the next day. By contrast, higher positive affect (node 4) predicted higher next-day perceived Challenge progress for you.

Not all the arrows and loops are explained here, but we encourage you to take these examples to go node by node and look at each of your associations across your Learning Challenge! Here are some questions to help you reflect on your network graph:

- Based on what we know about the arrows, can you identify your *strongest predictor* of perceived Challenge progress? Among all the relationships for perceived Challenge progress, which make the most sense to you?
- Do you think this graph is a good way to understand how these five variables interacted for you during your Learning Challenge? What information would you add to (or take away from) this graph if you were the one creating Research Summaries for Challengers?
- Imagine that you are going to share this graph with a friend or family member. How would you describe it to them? What would be the things you would point out first and why? What questions do you think they would have?
- What is the most surprising association on your network graph? What is the least surprising?
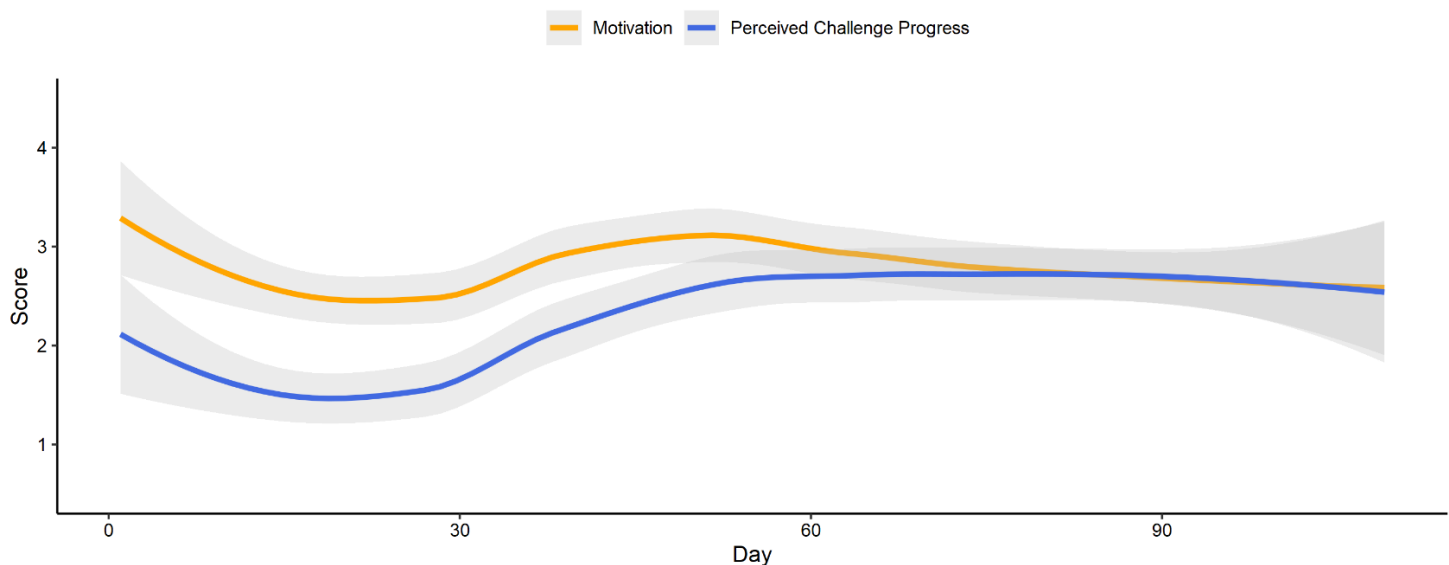
Your Daily **Perceived Challenge Progress** and **Motivation**

Your thickest line in your whole network was between your perceived Challenge progress and motivation. This means that your motivation on any given day was very predictive of how much progress you felt you made in your Challenge that day. On a day when you felt like you were making a lot of progress in your Challenge, you tended to feel more motivated. Similarly, on days with lower perceived Challenge progress, you tended to report lower motivation. Let's take a closer look at these two variables and how they changed throughout your Learning Challenge.

Your Daily Perceived Challenge Progress and Motivation

Below is a graph that shows how your perceived Challenge progress and motivation changed during your Learning Challenge. The x-axis (the numbers running horizontally) is the day of your Learning Challenge. The y-axis (the numbers running vertically on the left) shows all the possible scores of perceived Challenge progress or motivation. In this study, perceived Challenge progress and motivation were both rated on a scale that ranged from 1 to 5, with higher scores indicating greater perceived Challenge progress or motivation. The lines running through the graph show the overall trends of your perceived Challenge progress (blue) and motivation (orange) throughout your Learning Challenge.

Your Perceived Challenge Progress and Motivation by Day

Summary of Your Daily Perceived Challenge Progress and Motivation

Both your perceived Challenge progress and motivation seemed to stay relatively stable throughout your Learning Challenge. Your motivation started out a little higher than your perceived Challenge progress, but both of them ended close to the midpoint of the scale. Just as we would expect, given the strong relationship between the two, your perceived Challenge progress and motivation follow a very similar trend throughout your Challenge. Here are a few questions for you to reflect on as you look over your graph.

- As we saw in the network graph, the graph here shows a strong relationship between progress and motivation. How might perceived Challenge progress and motivation influence each other? Is this tight coupling true for all people?
- The gap between your motivation and your perceived Challenge progress closed as your Learning Challenge went on. Did you feel this change happening in the moment?

Your Daily Positive and Negative Affect

Just like your network graph showed your perceived Challenge progress and motivation were related, it also showed that your positive and negative affect seemed to be strongly related to each other. However, instead of a positive relationship, your positive affect was negatively related to your negative affect. In other words, higher positive affect was associated with lower negative affect on the same day, and higher negative affect was associated with lower positive affect on the same day. Let's take a look at your positive and negative affect throughout your Learning Challenge.

Your Positive and Negative Affect During Your Learning Challenge

Below is a graph that shows how your positive and negative affect changed during your Learning Challenge. The x-axis (the numbers running horizontally) is the day of your Learning Challenge. The y-axis (the numbers running vertically on the left) shows all the possible scores of positive and negative affect. In this study, positive and negative affect were both rated on a scale that ranged from 1 to 5, with higher scores indicating greater positive or negative affect. The lines running through the graph show the overall trends of your positive (blue) and negative (orange) affect throughout your Learning Challenge. Days 17–27 of your Challenge are particularly interesting, so the picture below the main graph "zooms in" on these days to see this trend more easily.
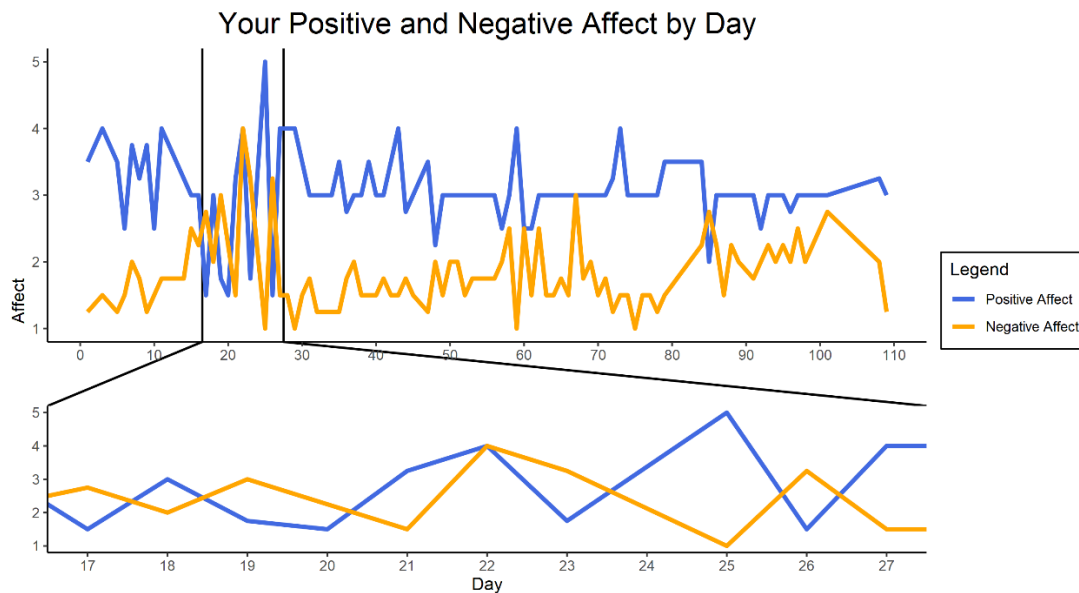
Summary of Your Positive and Negative Affect During Your Challenge

Both your positive and negative affect seemed to stay relatively stable throughout your Learning Challenge. This means that your positive affect could take on different values depending on the day but hovered around a 3 on average. Your negative affect also went up and down but seemed to stay around a 2 on average. For the most part, you can see that your positive affect generally stayed above your negative affect: you seemed to have very few days where you had more negative affect than positive affect. Here are a few questions for you to think about as you look over your graph.

- During days 17–27 of your Learning Challenge (which is shown on the zoomed-in part), your negative affect was sometimes above your positive affect. What was happening then? Were you facing any challenges, and what did you do to overcome them?
- Based on the three graphs in this Research Summary, how do you think your positive and negative affect might be related to your perceived Challenge progress, motivation, and life satisfaction? How might you be able to use this knowledge about yourself to help you develop greater awareness of your emotions and pursue your long-term goals after GripTape?

Code used to create the Research Summary

Loading, cleaning and editing data:

```r
# set working directory to the folder I have all my files
setwd("C:/Users/01csh/R-workspace/GripTape")

# load data
dat <- read.csv("lc16daily_9.15.csv", sep = ",", header = TRUE,
                stringsAsFactors = FALSE)

# download packages
library(tidyverse) # used for working with the data
library(ggplot2) # used for plotting

# convert variables to numeric
dat[c("Day", "prog_1", "motiv_1", "purpose_1", "lifesat_1", "esteem_1",
      "clarity_1", "em_1", "em_2", "em_3", "selfreg_1", "selfreg_2",
      "selfreg_3", "selfreg_4", "affect_1", "affect_2", "affect_3", "affect_4",
      "affect_5", "affect_6", "affect_7", "affect_8")] <-
  sapply(dat[c("Day", "prog_1", "motiv_1", "purpose_1", "lifesat_1",
               "esteem_1", "clarity_1", "em_1", "em_2", "em_3", "selfreg_1",
               "selfreg_2", "selfreg_3", "selfreg_4", "affect_1", "affect_2",
               "affect_3", "affect_4", "affect_5", "affect_6", "affect_7",
               "affect_8")],
         as.numeric)

# create new rows with the averages for environmental mastery, self
# regulation, and positive and negative affect
dat <- dat %>%
  dplyr::mutate(em = rowMeans(.[,c("em_1", "em_2", "em_3")], na.rm = TRUE),
                selfreg = rowMeans(.[,c("selfreg_1", "selfreg_2", "selfreg_3",
                                        "selfreg_4")], na.rm = TRUE),
                pa = rowMeans(.[,c("affect_1", "affect_2", "affect_3",
                                   "affect_4")], na.rm = TRUE),
                na = rowMeans(.[,c("affect_5", "affect_6", "affect_7",
                                   "affect_8")], na.rm = TRUE)) %>%
# rename variables
  dplyr::rename(prog = prog_1,
                motiv = motiv_1,
                purpose = purpose_1,
                lifesat = lifesat_1,
                esteem = esteem_1,
                clarity = clarity_1)
```

In a behind-the-scenes step (for the sake of data privacy), I pull all of the individual Challenger's data, removing rows of NAs, into its own data frame called "dat2".

Plot 1: Unified Structural Equation Model

This network graph is based on a machine learning tool called a unified structural equation model (uSEM), which explores the associations between a set of the Challenger's variables (I chose 5) during their Learning

Challenge. It represents the valence (positive or negative), strength, direction, and temporality (same-day or next-day) of relationships between each variable, as well as each variable's stability within itself. I explain this network and how to interpret it in more detail on the second and third content pages of the Research Summary.

```r
# the package "pompom" is used to implement the uSEM and plot the network graph
library(pompom)

# pick the variables I want to include in the model: perceived Challenge
# progress, motivation, life satisfaction, positive affect, and negative affect
# and put them into a vector that will be used later to fit and plot the model
vars <- c("prog","motiv","lifesat","pa","na")

# subset to the variables you want
dat_vars <- dat2 %>%
  dplyr::select("CID", "Day", all_of(vars))

# z-standardize the variables, and replace the original values with the
# newly standardized values
dat_z <- dat_vars
for(var in vars) {
  dat_z[var] <- dat_vars[var] %>%
    sapply(function(x) c(scale(x, center = TRUE, scale = TRUE)))
}

# fit a unified structural equation model (uSEM).
usem <- uSEM(var.number = length(vars), # number of variables in the network
             data = dat_z[ , c(vars)], # columns of the standardized variables
             lag.order = 1, # observations lagged by 1 day
             verbose = FALSE, # verbose and trim cut down on the text we see
             trim = FALSE)

# save the model summary from the usem, needed for plotting later.
ms <- model_summary(model.fit = usem,
                    var.number = length(vars),
                    lag.order = 1)

# this prints the answer to the question whether 3 of 4 fit indices of the model
# pass acceptable thresholds. this must return TRUE, or else the path estimates
# are inaccurate, the model does not fit the data, I'm not reporting
# something reliable, and likely something is wrong with my data inputs. USEM is
# a machine learning tool that typically runs until it fits the data well (i.e.,
# results in "TRUE").
print(sum(ms$cfi > .95,
          ms$tli > .95,
          ms$rmsea < .08,
          ms$srmr < .08) >= 3)

# plot the network and save it as a png:
png("network.png", type = "cairo")
# this "opens" a png file and sets the type to cairo, which removes aliasing
# to make the lines in the plot smooth in the final output
plot_network_graph(ms$beta, var.number = length(vars))
# plot the graph "onto" the png. this uses the function plot_network_graph()
```

```
# from the package "pompom", and uses the beta matrix from the uSEM model
# summary that I saved from earlier, which contains all the correlations
# between the variables, same-day and next-day.
dev.off()
# "close" the png. after this, the png will be saved in my working directory
# under the name I gave it ("network.png")
```

Plot 2: Twin Loess Line Plot of Perceived Challenge Progress and Motivation

```
# to create this plot, I had to pivot the data to longer form (less columns,
# more rows), in order to be able to subset the data by the variable
# (perceived Challenge progress and motivation)
dat2_long <- dat2 %>%
  pivot_longer(c("prog","motiv"), names_to="var", values_to="level")

progmotiv <- ggplot(dat2_long, aes(Day, level, color = var)) +
  # maps the color of the line to the variable, creating 2 different-colored
  # lines for perceived Challenge progress and motivation
  geom_smooth(method=loess, size=1.3, se=T, fill = "gray", alpha = .3) +
  # draws the loess lines
  scale_color_manual(values=c("orange", "royalblue"),
                     labels=c("Motivation", "Perceived Challenge Progress")) +
  # sets the colors and labels for each line
  scale_y_continuous(limits=c(.5,4.5), breaks=c(1:5)) +
  #edit the y axis for better readability and visuals
  theme_classic() +
  labs(x = "Day",
       y = "Score",
       color = "" , # makes the legend title blank
       title = "Your Perceived Challenge Progress and Motivation by Day") +
  # set labels for the axes and graph
  theme(plot.title = element_text(hjust=0.5, color="black", size=rel(1.7)),
        axis.title = element_text(color="black", hjust=0.5),
        axis.text=element_text(color="black"),
        aspect.ratio=5/16,
        legend.position = "top"
  ) # other text and scale customizations

# saves the graph as a png, with "cairo" to make lines smoother just as above
ggsave("progmotiv.png", type = "cairo", scale=1.5, plot = progmotiv)
```

Plot 3: Twin Line Plot of Positive and Negative Affect

```
library(ggforce)
colors <- c("Positive Affect" = "royalblue", "Negative Affect" = "orange")

pana <- (ggplot(dat2, aes(x = Day)) +
    geom_line(aes(y = pa, color = "Positive Affect"), size = 1.3) +
    geom_line(aes(y = na, color = "Negative Affect"), size = 1.3) +
```

```
    scale_color_manual(values = colors) +
    facet_zoom(dplyr::between(Day, 17, 27), horizontal = TRUE, zoom.size=0.5) +
    scale_x_continuous(breaks = function(Day)pretty(Day, n = 10)) +
    theme_classic() +
    labs(x = "Day",
         y = "Affect",
         color = "Legend",
         title = "Your Positive and Negative Affect by Day") +
    theme(plot.title = element_text(hjust=0.5,size = rel(1.7)),
          legend.position= c(1.1, 0.5),
          legend.background = element_rect(size=0.5, linetype="solid", colour = "black"),
          plot.margin=unit(c(.75,5.5,.75,.75),"cm"),
          aspect.ratio=5/16))

# save as png with cairo
ggsave("pana.png", type = "cairo", scale = 1.5, width=7, height = 3.8,
       plot = pana)
```