Box and Whisker Plots and Projecting Dates: How Box and Whisker Plots Can be Used to Create a More Accurate Period Tracker for Women

Vanessa Rowlin
PSYC6135
Michael Friendly
York University
April 4 2019

Box and Whisker Plots and Projecting Dates: How Box and Whisker Plots Can be Used to Create a More Accurate Period Tracker for Women

Mobile-Health Technology & Data Visualization

From large organizational bodies, such as companies or governments, to as small as a single individual, we all employ data in some way to understand, plan and take action (McCandless, 2010; Munzner, 2014). In line with this, despite being a relatively new phenomenon, people are increasingly engaging with mobile health applications for self-monitoring various areas of their health; anything from tracking their steps to monitoring their sleep patterns (Anderson, Buford & Emmerton, 2016). With its constant presence and connectivity, mobile and smart phone technology easily integrate into one's life to help us effectively manage everyday tasks (Chen et al., 2012).

One prominent producer of mobile health applications would be Apple Inc. Based on the statistics from mHealth Initiative Inc., the number of mobile health applications on the Apple Inc. mobile platform (iOS) was greater than the numbers of applications on any other platforms by the end of 2009, and Apple Inc. even led in the application market in the year of 2010 too (Liu et al., 2011). Thus, as Apple Inc. is at the forefront of mobile health applications, they need to have an awareness of the best visualizations to support their users. Unfortunately, as mobile health applications are a recent phenomenon, proven guidelines on what works in mobile health app development have not emerged yet (Liu et al., 2011). As many of these mobile health applications are based in self-care, the responsibility to use them rests upon the consumer; that being said, these mobile health applications need to be designed in a way that allows users to make the best next steps for their health (Anderson, Buford & Emmerton, 2016).

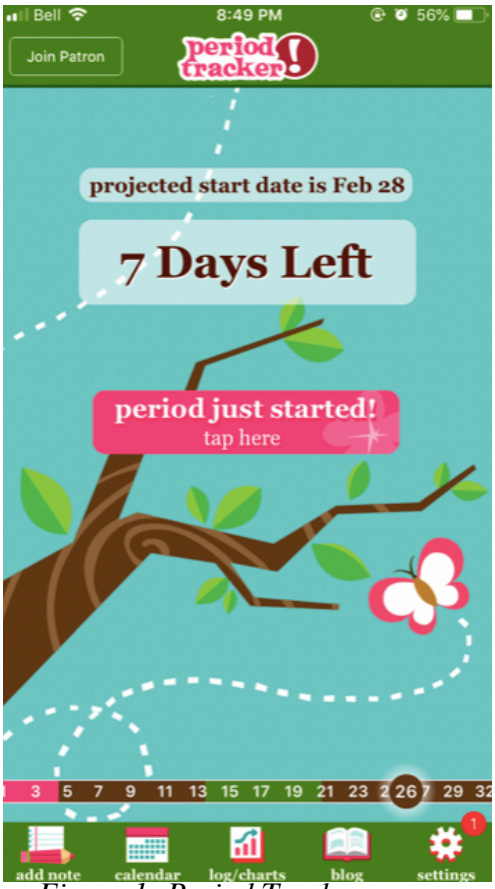


Figure 1: Period Tracker.

This brings me to the specific mobile health application in question for this research proposal; in particular, this paper will be critically analyzing where a common Apple Inc. application called Period Tracker by GP Apps has gone wrong when designing the most optimal visualization for its data, and therefore how it has failed to provide the consumer the best advice for self-care during their menstrual cycles. This paper will also recommend a different visualization that would prove more helpful for developing a consumer's understanding and helping them to make informed decisions about their menstrual cycle.

Data Visualization Features, Working Memory & Decision-Making

When produced properly, visual representations of data can make it easier to see patterns and make certain information more salient. As humans, we are naturally drawn to images, so much so that it actively takes effort from working memory to shift our focus away from them when they are in our visual field (Sanchez & Wiley, 2006**).** Therefore, mobile health applications, such as Period Tracker, could best take advantage of this natural inclination by developing the visual components of their applications in a way that best serves their purpose.

In consensus with graphic designers, data scientists and academics alike, there is the general idea that the most optimal visualizations should be presented in the clearest possible way in order to increase both understanding of the content being shown, as well as assist the user to make the best-informed decision (Tufte, 2001; Cairo, 2013).

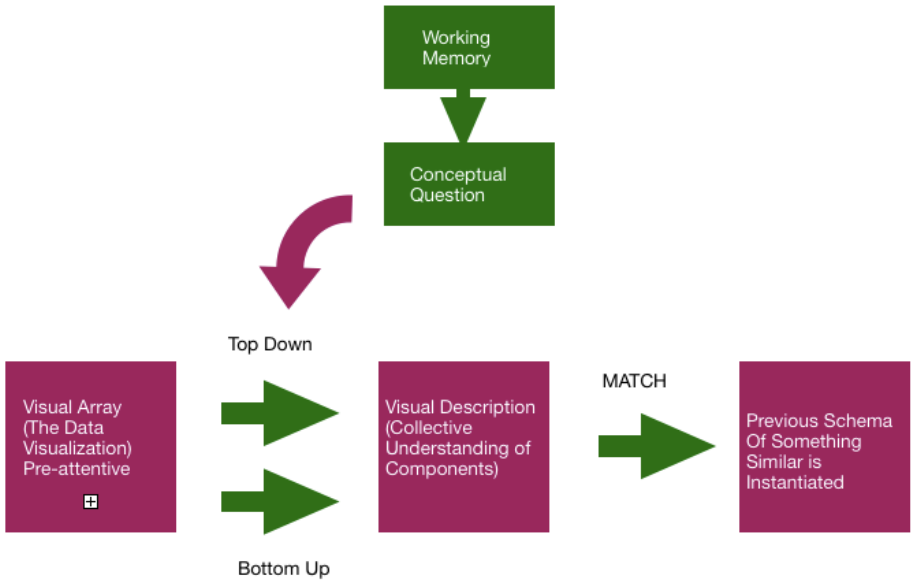


Figure 2: Padilla and colleagues (2018) model for Decision Making from Data Visualizations.

According to Padilla and colleagues (2018), when a user encounters a data visualization they engage in both top down and bottom up processes. They are taking in the pre-attentive graphic features, but also mentally searching their long-term memory for relevant knowledge (a graphic schema) for interpreting the data visualization. Utilizing pre-attentive graphic features is important in data visualizations because according to visual research, pre-attentive graphic features (e.g. line orientation, width, length and colour) are processed with little effort, enabling the processing of more data without overloading the decision maker (Lurie and Mason, 2007). Next, through a sort of match analysis, they will try to find one that is most similar to the visual array and when it does, they will be ready to ask concept questions such as “what is this graph trying to say?” When there is a mismatch between the task and visualization, the additional transformation accounts for the increased time taken to make their decision because additional work is needed by working memory (Aigner et al., 2011). The reason for this is because the mind of the viewer needs to understand and retain the information through bottom- up processing

while trying to re-configure the visualization to best match the appropriate graphic schema from long term memory. To simplify complex ideas in a visual representation, it is often helpful to keep distracting or irrelevant information to a minimum (Hullman & Diakopoulos, 2011**).** When graphic visualizations in mobile health applications use graphics that are overly complicated or do not match up with the message they are trying to convey, it adds unnecessary work for the user's working memory and makes the decision-making process all the more complicated.

While there are many different types of visual features that can be presented in mobile health applications (e.g. images of people engaging in an action, text reminders, calendars, etc), this paper is primarily interested in how the presentation of data inputted into the mobile health application by the user can be portrayed graphically in the most effective manner for the user's understanding. This paper will be analyzing the Period Tracker application and argue against their choice of a histogram for presenting past and future menstrual cycles, while also recommending a box and whisker plot. This modification in the mobile health application will therefore assist women in making more informed decisions every month without fear of being inconvenienced by the start date of their period.

Current Problems with Period Tracker's Data Visualization

Period Track by GP Apps is a women's health diary with a popular rating of 4.8 out of 5 stars with a 2.6K rating. Amongst its various features, Period Tracker logs the dates in which the user's period has started and stopped, as well as calculates the average of the user's past 3 months' menstrual cycles in order to predict the start date of their next period. The user may also view their past period dates in a histogram format, as well as the number of days each period if they want to see general trends in either earlier or later period start dates.

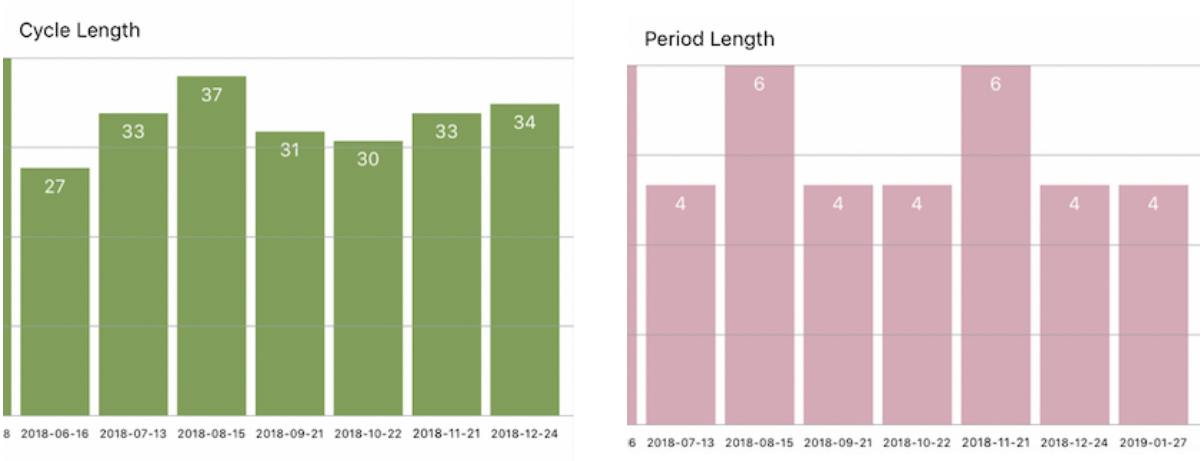


Figure 3: Histograms of Cycle Length and Period Length in Period Tracker.

To start off, one area of concern with the Period Tracker is how it ineffectively uses the number of days in each cycle inputted into the mobile application and creates an inaccurate measure of future period start dates. For instance, one month may have a number of 28 days for the cycle, but another may have as many as 39 days. For a typical woman their menstrual cycle can change from month to month due to a variety of factors such as stress, amount of daily sleep, personality factors, age, nutrient intake, oral contraceptives, intense exercise, weight fluctuations,

giving birth, menopause or various diseases (Dasharathy et al, 2012; Treloar et al, 1967; Hillard, 2014; Tirelli, Cagnacci & Volpe, 2008). However, the problem with the Period Tracker lies in the fact that it will only take the average of the past three months and project the future period date. Over time, with constant use of this application, the Period Tracker has more than three points of data in which it could use to find a more accurate indicator of a future period start date. Instead of only using the average of three points, it could instead use as many points as is made available in order to provide a more accurate picture of one's cycle. In addition to providing a misconstrued view of the data, by only using three points of data, the projected date given by the Period Tracker may be influenced by outliers in the user's menstrual cycles, which could be brought about by various factors (i.e. stress, nutrition, exercise, sleep). Analyses and graphics that do not involve identifying outliers can often lead to inaccurate conclusions (Hyndman et al, 2010). Thus, this poor use of data impacts the ability to provide a more useful visualization for a user to make adequate decisions regarding their period start date.

Similarly, in addition to the problematic use of averaging only three data points and their susceptibility to outliers, the Period Tracker application's use of a histogram to visually display the number of days in one's cycle is incredibly problematic for the user if they want to see general trends in either earlier or later period start dates. Typically, a histogram is an excellent graph to use to examine patterns and see how data changes over time, however this is not an appropriate data visualization when the mobile health application primary objective is to project an appropriate date for a women's period. Over time as a user inputs more cycles into the mobile application, they will have more data points than one can find on the screen. This means they will have to slide the screen to the left and see those from the beginning.

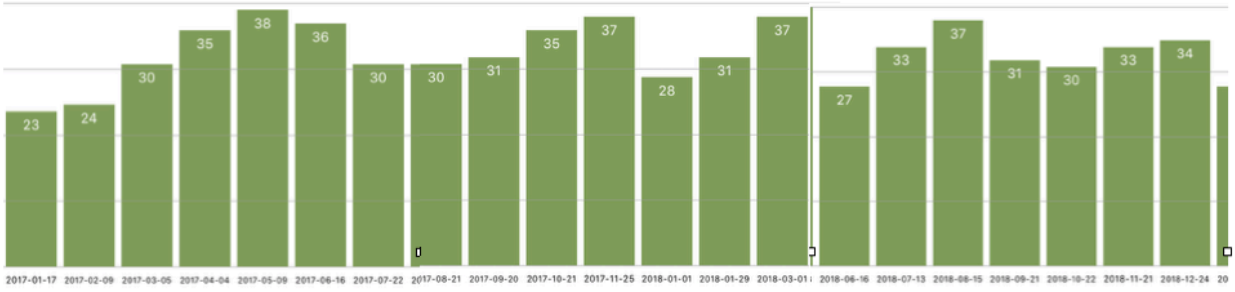


Figure 4: Example of an over-extended Histogram for Cycle Length in Period Tracker for long time users.

As a visual itself, an over-extended histogram, like the one in Figure 4 above, is visually difficult to analyze for a user, even with the scrolling option made available by the application. In a study by Haroz and colleagues (2015), they found that error levels in graphic interpretation increase from left to right for complex graphics. Haroz and colleagues outline that the visual system can encode small amounts of items, with a maximum of 5. As the number of items increase beyond that range, the visual system is forced to engage in a noisy estimation of what the visual graphic is supposed to depict. According to Padilla and colleagues' (2018) model of decision-making, users would put additional load on their working memory as they scan and hold the abundance of visual information in their mind .When you present your data in a succinct

manner, you allow their brains to process comparison, patterns and differences more rapidly (Few, 2014; Koch et al, 2006).

According to David McCandless, data journalist and author of the website and book *Information is Beautiful,* a critical factor to appropriately presenting one's data into a readable format includes compressing huge amounts of information into coherent small frames (McCandless, 2010; Few, 2014). In this respect, the Period Tracker fails in utilizing the histogram appropriately and instead overloads the user with graphical data far too complex and distracts them from the core message of the data and the application in general (Tufte, 2011).

Recommend Change: Box Plots as an Alternative Data Visualization

It is understandable why the programmers may have not utilized a box-plot design for their Period Tracker application. Generally, box and whisker plots are not available in non-statistical software tools, such as Excel (Spitzer, 2014). In fact, of the top seven statistical graphics used, bar graphs and histograms are in the top three (Taylor, 2018). Box and whisker plots do not even make the list in most commonly used graphical visualizations. In this respect, one may surmise that their visualization choices reflect more so common trends in graphical culture, however the box and whisker plot is still a superior choice for the use of a Period Tracker.

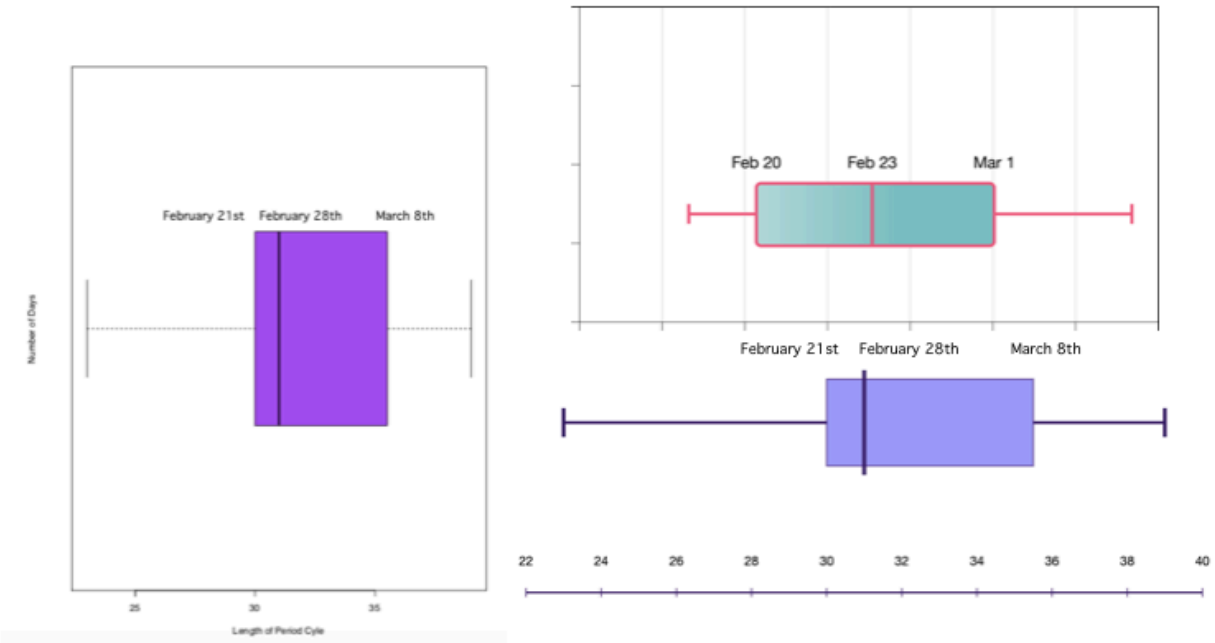


Figure 5: Examples of Box and Whisker Plots that could be used in Period Tracker. Left: Made in R, Top Right: Made in Photoshop, Bottom Right: Made in MetaChart.

A box and whisker plot is good choice for this application because it will present their median number as well as the two quartiles (first and third quartile) that provide the lower and upper range of the project start of the period date. By utilizing a box and whisker plot, a woman could see a visual similar to Figure 5. A woman would have the estimated projected start date,

but would also have an earlier date (quartile one) as well as a later date (quartile three) that are estimates based upon confidence intervals from the data set. Furthermore, the text written above the box and whisker plot will also benefit the user's use of the visual for their decision making. According to Haroz and colleagues (2015), "text labels may be recognized more quickly, as a lifetime of reading renders recognition of single words a surprisingly fast and automatic process" (Friendly, 2019, lecture 6).

In addition, as a graphic visual, a box and whisker plot is an excellent way to summarize large amounts of data in one clear graphic (Reese, 2005). As the user inputs more information into Period Tracker, the box and whisker plot will be able to handle the extra data without disturbing the graphic visually and making it more visually difficult for the user to understand. For instance, the impression of variability in a box and whisker plot can show outliers without being inherently impacted by them. While a mean is an easily calculable number for projection, it can be impacted by larger or smaller numbers. The median does not fall victim to the influence of outliers and instead is simply the middle number of all data points entered by the user. When considering this as an alternative, it is imperative to realize that sometimes outliers are important to visualizing the full story of the data, however one must first understand the type of data one is analyzing in the first place. In particular, the data that would be inputted into the Period Tracker will sometimes include abnormal period cycles (Dasharathy et al., 2012).

Overall, benefits of this format in terms of data visualization is that it is presented in the clearest possible format and that task specific to knowing the date range for the start of one's period. In line with Padilla and colleagues (2018) working memory and decision-making model for visualizations, this would reduce the load on working memory when the user is using top-down processes and ask the conceptual question of "what is this graph trying to say?". In this respect, the reduction in cognitive load would not only reduce the level of difficulty in the decision-making process, but also the time it takes to make informed choices about what women need to do to prepare for the start of their period date.

Implications of the New Data Visualization

Improving the Period Tracker can produce considerable benefits for those female users who use the mobile application. By replacing an over-extended histogram with a simple box plot with whiskers, as well as including all data points to find a median instead of an average, you would provide women with an easier visualization that would assist in many aspects of menstrual cycle.

Firstly, by including all data entered into the mobile application, instead of merely three points, the application will be able to draw from more data, and therefore be able to provide a better indicator of what is most typical for the woman's menstrual cycle. That being said, this will mean that, projected date that is far more accurate and less hindered by those few months where she was, for instance, under considerable stress and therefore produced an outlier in her data set (Dasharathy et al., 2012). Even if there are extreme data points (outliers), over time, and with the collection of more data, these will impact the mobile application ability to project a date less and less. While this may seem like a minor issue, by having a more accurate start date, a woman can not only have the tools ready, but also make pro-active decisions in the weeks leading up to it.

In terms of the more pro-active decisions a woman can make with a more accurate start date, they can range anything from skin care, diet, even physical activity. For instance, according to Shah and Maibach (2001), as one gets closer to closer to the start of their period, changes will occur in the levels of estrogen, testosterone and progesterone that can lead to dry skin. As one's skin gets less firm and drier, it opens up the possibility for micro abrasions that, when filled with oil, can lead to break outs. In this respect, if a woman has a more accurate start date, they can reduce the chance of hormonal acne with proper hygiene appropriate for the time (hydration). In one way tied to proper facial hygiene, women can also make adjustments to their schedule to reduce menstrual cramps and pain. She will be able to schedule her dates appropriately in so that, during dates when she is within the box plotted dates, she could refrain from high stress activities, which would increase menstrual pain (Nohara et al, 2010; Yamato et al., 2009). She could also engage in activities that reduce menstrual pain (e.g. exercise, reducing sugar intake) (Wulandari &Afiliana 2017). Lastly, she would make sure to have the required equipment (e.g. appropriate undergarments and menstrual padding) in her purse and around the house. In this respect, the appropriate visualization of data would vastly assist her life not only during this week but in general, as she could plan around the five days.

References

Aigner, W., Miksch, S., Schumann, H., & Tominski, C. (2011). *Visualization of time-oriented data*. Springer Science & Business Media.

Anderson, K., Burford, O., & Emmerton, L. (2016). Mobile health apps to facilitate self-care: a qualitative study of user experiences. *PLoS One*, *11*(5), e0156164.

Cairo, A., 2013. The Functional Art. New Riders, Berkeley, CA

Chen, C., Haddad, D., Selsky, J., Hoffman, J. E., Kravitz, R. L., Estrin, D. E., & Sim, I. (2012). Making sense of mobile health data: an open architecture to improve individual-and population-level health. *Journal of medical Internet research*, 14(4).

Dasharathy, S. S., Mumford, S. L., Pollack, A. Z., Perkins, N. J., Mattison, D. R., Wactawski-Wende, J., & Schisterman, E. F. (2012). Menstrual bleeding patterns among regularly menstruating women. *American journal of epidemiology*, *175*(6), 536-545.

Few, S. (2014a). 'Data Visualization for Human Perception.' *The Encyclopedia of Human-−−Computer Interaction*, 2nd edn. <https://www.interaction-−− design.org/encyclopedia/data_visualization_for_human_perception.html > [Accessed Feb. 2015].

Friendly, M. (2019). Graphical Perception [Powerpoint slides]. Retrieved from http://euclid.psych.yorku.ca/www/psy6135/

Haroz, S., Kosara, R., & Franconeri, S. L. (2015, April). Isotype visualization: Working memory, performance, and engagement with pictographs. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems* (pp. 1191-1200). ACM.

Hillard, P. J. A. (2014). Menstruation in adolescents: what do we know? And what do we do with the information?. *Journal of pediatric and adolescent gynecology*, *27*(6), 309-319.

Hullman, J., & Diakopoulos, N. (2011). Visualization rhetoric: Framing effects in narrative visualization. *IEEE transactions on visualization and computer graphics*, *17*(12), 2231-2240.

Hyndman, R. J., & Shang, H. L. (2010). Rainbow plots, bagplots, and boxplots for functional data. *Journal of Computational and Graphical Statistics*, *19*(1), 29-45.

Koch, K., McLean, J., Segev, R., Freed, M., Berry, M., Balasubramanian, V., and Sterling, P. (2006). 'How Much the Eye Tells the Brain', *Current Biology,* 16(14): 1428–34.

Liu, C., Zhu, Q., Holroyd, K. A., & Seng, E. K. (2011). Status and trends of mobile-health applications for iOS devices: A developer's perspective. *Journal of Systems and Software,* 84(11), 2022-2033.

Lurie, N. H., & Mason, C. H. (2007). Visual representation: Implications for decision making. *Journal of marketing*, 71(1), 160-177.

McCandless, D. (2010). 'The Beauty of Data Visualization.' <www.ted.com/talks/david_mccandless_the_beauty_of_data_visualizati on?language=en>

Munzner, T., 2014. Visualization Analysis & Design. CRC Press, Boca Raton, Florida.

Nohara, M., Momoeda, M., Kubota, T., & Nakabayashi, M. (2010). Menstrual cycle and menstrual pain problems and related risk factors among Japanese female workers. *Industrial Health*, 1012100044-1012100044.

Padilla, L. M., Creem-Regehr, S. H., Hegarty, M., & Stefanucci, J. K. (2018). Decision making with visualizations: a cognitive framework across disciplines. *Cognitive research: principles and implications*, *3*(1), 29.

Reese, R. A. (2005). Boxplots. *Significance*, *2*(3), 134-135.

Sanchez, C. A., & Wiley, J. (2006). An examination of the seductive details effect in terms of working memory capacity. Memory & Cognition, 34, 344-355.

Shah, M. G., & Maibach, H. I. (2001). Estrogen and skin. *American journal of clinical dermatology*, *2*(3), 143-150.

Spitzer, M., Wildenhain, J., Rappsilber, J., & Tyers, M. (2014). BoxPlotR: a web tool for generation of box plots. *Nature methods*, *11*(2), 121.

Taylor, C. (2018). 7 Graphs Commonly Used in Statistics. [online] ThoughtCo. Available at https://www.thoughtco.com/frequently-used-statistics-graphs-4158380 [Accessed 16 Jan, 2019].

Tirelli, A., Cagnacci, A., & Volpe, A. (2008). Levonorgestrel administration in emergency contraception: bleeding pattern and pituitary-ovarian function. *Contraception*, *77*(5), 328-332.

Treloar, A. E., Boynton, R. E., Behn, B. G., & Brown, B. W. (1967). Variation of the human menstrual cycle through reproductive life. *International journal of fertility*, *12*(1 Pt 2), 77-126.

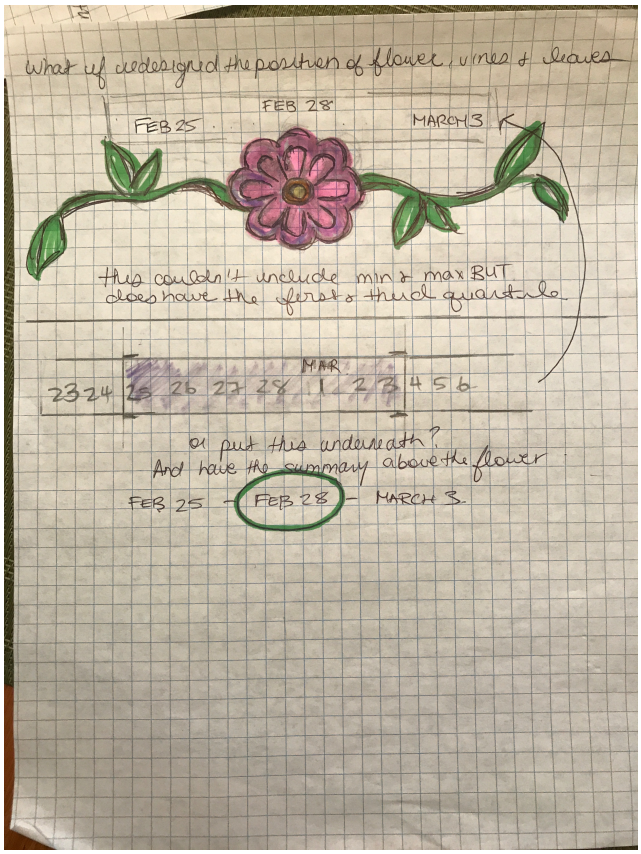
Tufte, E., 2001. The Visual Display of Quantitative Information, second ed. Graphics Press.

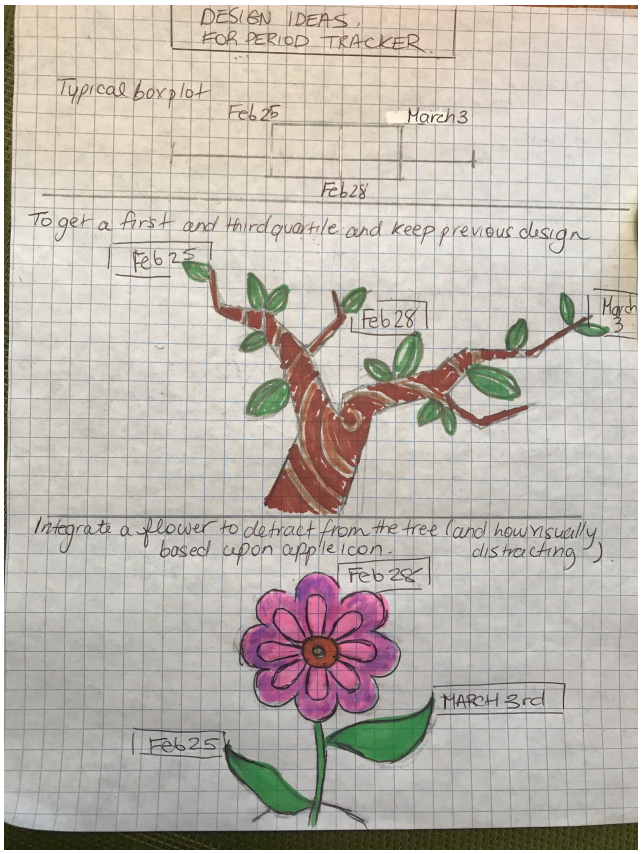
Wulandari, S., & Afriliana, F. D. (2017). The Effect of Consumption Dark Chocolate Against Primary Dismenore Pain for Girls Teenager at Kediri 5 High School. . *Indian Journal of Medical Research and Pharmaceutical Sciences*, *4*(2).

Yamamoto, K., Okazaki, A., Sakamoto, Y., & Funatsu, M. (2009). The relationship between premenstrual symptoms, menstrual pain, irregular menstrual cycles, and psychosocial stress among Japanese college students. *Journal of Physiological Anthropology*, *28*(3), 129-136.

Appendix 1: Visual Design Ideas for Optimal Box and Whisker Plot for Period Tracker

a





Appendix 2: Example Data set of Cycle Days in Period Tracker

Cycle Number	Days Per Cycle
1	25
2	24
3	30
4	35
5	38
6	36
7	30
8	30
9	31
10	35
11	37
12	28
13	31
14	37
15	31
16	39
17	27
18	33
19	37
20	31
21	30
22	33
23	34
24	27

Appendix 3: R Script for Menstrual Cycle Length Box Plot in R

```
> MenstrualCycleLength=c (23, 24, 30, 35, 38, 36, 30, 30, 31, 35, 37, 28, 31, 37, 31, 39, 27, 33, 37, 31, 30, 33, 34, 27)
> median(MenstrualCycleLength)
[1] 31
> boxplot(MenstrualCycleLength, horizontal = TRUE, xlab= "Length of Period Cycle", ylab= "Number of Days", col = c("purple"))
```