Response to reviewers

JCGS-20-344 21 April 2021

We would really like to thank the editor and the reviewers for their encouragement and detailed feedback which has helped to improve the paper in terms of both concision and precision as well as reaching out to more readers. The point by point description of changes are below:

Reviewer 1 comments

Q1. I am not convinced that it is required to introduce the concept of grammar of graphics and embed the used visualization in this framework on such a level of detail like in Section 5. Of course some elements of the grammar of graphics are used in Section 5.2 to argue about facets, levels, aesthetic variables, and mappings, but if the reader does not know about this framework, the introduction and explanations about this concept/framework provided in Section 5 is not enough to follow the later argumentation. I would like to see some more details on the concept/framework on the grammar of graphics, or not using the concepts in the parts of Section 5.2 and paraphrasing it without using the terms from the grammar of graphics. I think that it is doable to explain the construction of the visualization without using this concept/framework, because the mappings are quite straight forward. I would like to see (a) either extend the introduction to grammar of graphics and the basics needed to argue in more detail about the embedding of your visualisations using this concept/framework, or (b) remove it.

A1. removed.

Q2. It is not clear from the paragraph in Section 5.1 what basic plot choices are implemented and used and supported for visualisation. There is only the list of plots to display descriptive statistics provided. It would be important to describe which are implemented and possible to use. The later examples use primarily letter-value plots, then also box-plots, violin plots and line charts with area quantile plots. The letter-value plots are used in major examples of the paper, therefore I would argue to provide some more details on this visual representation and what the 5 colours and letters mean (e.g. LV=letter value, M = median, F = fourths, E = eighths, D = sixteenths, C = thirty-seconds), as it is not as widely known like standard box-plots. This would help greatly in better following the argumentation and explanations.

A2. added in Section 5.1

Q3. I have to confess that I do not know anything about the sport cricket and it may be unthinkable for people from countries where cricket is a national sport, but I really had difficulties to follow the example in Section 6.2. The basic hierarchical concept of over—inning–match–season was understandable, but I think that some parts need to be revisited and clarified for also reaching readers not knowledgeable in cricket. Especially the discussion on the interesting questions raised are a bit difficult to follow without knowing much about cricket. For example some terms should be at least briefly explained or removed if not necessary for the explanation, e.g., Twenty20 format, fielding, bowling, dismissal, wicket. I do know a lot about different sports and also the basic idea and structure in many sports, but cricket was never on my radar. Apologise my ignorance about cricket.

A3. Thanks for the advice. We have added a description of the game cricket in the first paragraph of Section 6.2 so that each new terminology applicable for cricket (and relevant for the following analysis) is highlighted in italics and briefly explained. We believe that this will help the readers not knowledgeable in cricket to also follow the presented analysis.

1 Reviewer 2 comments

1.1 Notation/Definition clarification

Q1.What is the definition of 'period'? Is the period defined on the grouping or on the granularity? Is the period denoted by P or (R, P) in Definition 6? From Pages 6 Line 55 to Page 7 Line 7, period was referred to an integer (7), and two time granularities (one year and 400 years). Are they consistent?

A1. Period is defined for a pair of granularities (G, H). In Definition 6, the period is defined by P, which is briefly defined while explaining Definition 6. We have now made the notations consistent.

Q2. What is the definition of 'grouping'? Though 'G groups into H' was defined in Definition 5, there is not a notation of grouping (G, H) until we see grouping $(day, month)$ at Page 7 Line 7. Please rephrase the sentence each month is a grouping of the same number of days over years on Page 7 Line 5.

A2. We have used the term 'pair' instead 'grouping' to avoid conflict with the definition of 'groups into' and rephrased the sentence.

Q3. What is the definition of aperiodic linear granularities? On Page 11, Definition 10, M_i are aperiodic linear granularities. If M_i are aperiodic, then is M also an aperiodic linear granularity? What is the relationship between aperiodic linear granularity and aperiodic cyclic granularity?

A3. Aperiodic linear granularities are those that can not be defined by Definition 6. If M_i is periodic, then $M = \bigcup M_i$ is periodic, but if each M_i's are aperiodic, then M could be periodic or aperiodic. In Definition 10, we want M to be aperiodic. Definition 10 and introduction of Section 3.3 briefly rephrased to improve clarity.

Q4. w in T_w is not well defined for circular granularities in Section 3.2 & 3.5. The definition of T_w uses T_w itself in the formula making it hard to understand. Relevantly, what is the interpretation of k where w = 0; 1; :::; k-1? In my opinion, the number of cycles should be defined for cyclic granularities, for example, $w_{B,G}(z) = \lfloor z/P(B;G) \rfloor$, which represent the cycle that the granule is in.

A4. R' is the number of granules of G' in each cycle. If z belongs to k^{th} granule of G', then $z \in T_k$. Here w mod R' represents the index that must be shifted to obtain $G'(w)$. The idea here is if we know the composition of each of the granules of G' in terms of granules of B for one cycle, we can find the composition of any granule of G' beyond a cycle since the 'pattern' repeats itself along the time domain due to the periodic property. The periodic property also ensures that $|T_w| = |T_{w \mod R'}|$ since every w^{th} and $(w + R')^{\text{th}}$ granule of G' will have the same number of granules of B. The term $\sum_{w=0}^{k-1} |T_w|$ denotes the number of granules of B till the $(k-1)^{\text{th}}$ granule of G'. Since $|T_w| = |T_{w \mod R'}|$, the number of granules of B till the $(k-1)^{\text{th}}$ granule of G' becomes $\sum_{w=0}^{k-1} |T_{w \mod R'}|$ in Definition 9.

1.2 Places that need clarification

Q1. Page 7, Definition 6. Is 'the number of granules of H' finite or infinite? If each granule $G(i)$ has the index $i \in Z_{>=0}$, then the phrase is similar as 'the number of non-negative integers.'

A1. Number of granules of H should be finite.

Q2. Page 7, Lines 19-20. It is claimed that quasi-periodic relationship has a finite number of periods where granularities behave in an anomalous way. What does the finite number of periods mean and why is it finite? Does it mean a finite number of granularity groupings? Does it mean a finite number of irregular cycles? By Definition 4 of Bettini & De Sibi (2000), $E_1, ..., E_z$ are the granularity exceptions, but the number of granularity exceptions does not have to be finite.

A2. 'Number of periods' rephrased as 'spans of time'. Bettini & De Sibi (2000) states in P57 L17 while introducing quasi-periodic relationships that " except for a finite number of spans of time where they behave in an anomalous way."

Q3. Page 8, Definition 8, what is the 'period' P(B,G)? Is it same as P in Definition 6? Is it a mapping or function?

A3. It is the same but stressed to make it clear that the period will be a 'function' of two granularities B and G.

Q4. Page 10, Line 14, Q4 week-of-month is probably not a good example for quasi-circular granularities. By Definition 4, week cannot be grouped into month, because there could exist $week(i) \subset month(j)$ when the i^{th} week crosses two months. Similarly, week is not finer than month by Definition 5. Since periodicity is defined on top of grouping, and circular is defined on top of periodicity, the scenario of Q4 cannot be defined. On the other hand, the period length for week-of-month is not an integer, but it is an integer for Q1 - Q3.

A4. removed

Q5. Page 10, Definition 9, is R' similar as the period P of the grouping (B,G')? Is R' a constant or variable with irregular mapping?

A5. R' is the number of granules of G' in the period P. For example, the grouping (day, month) has period 365 (ignoring leap years), then R = 12, since there are 12 months in 365 days. Given a period, R' is constant.

Q6. Page 11, Definition 10, what is the interpretation of n where i = 1, . . . , n?

A6. n is the number of Aperiodic linear granularities. Added in Definition 10.

Q7. Page 14, Lines 26-27. Is z the original index for the bottom granularity? If yes, why $C_{katun,baktun}(z) = \lfloor 1820z/20 \rfloor mod 20$? Shouldn't it be $zmod20$? If not, then what is z? By Definition 8, $z \in Z$ is the original time index based on the bottom granularity, but here z seems to be defined as the time index based on granularity katun.

A7. z is the original index for the bottom granularity *kin*. But the second line is replaced since the point is to show how the value of a multiple-order-up granularity can be obtained as a function of single-order-up granularities. It is sufficient to know the values of the single order granularities at z and not the exact value of z, so we chose to not expand it using z.

Q8. Page 11, lines 22-33. The explanation of Figure 3 is confusing. If the week-of-the- semester week type is a quasi-circular granularity, then the day-of-the-semester week type can be also quasi-circular granularity, as all days are nested in weeks. Actually the notation should not be 'week-of-the-semester week type', but 'semester week-of-the- semester week type'. The relationship between G and H should be clarified. It will be clearer if the authors can make a granularity 'semester day' (U) from B. Therefore we have the following granularities. $Q_{H,M}$ *: quasi-circular granularity, semester week -of-the-semester week type* $Q_{U,M}$: quasi-circular granularity, semester day-of-the-semester week type.

A8. done (to be)

Q9. Page 20, Line 47, 'for weekdays the interquartile range of consumption reduces over the year, whereas this pattern is not true for weekends' is hard to be justified. By Figure the IQR is given by the height of blue (LV=F) boxes. They are the largest in Quarter 2 and smallest in Quarter 4 for both weekdays and weekends.

A9. rephrased.

1.3 Other questions of interest.

Q1. Page 6, Definition 5, should S be a contiguous set of integers or not necessary?

A1. Not necessary.

Q2. Page 6, what is the difference between finer than and group into in Definitions 4 and 5? On Page 7, Line 34, they are used as alternative choices. Can you give an example that $G \preceq H$ does not imply $G \trianglelefteq H$, or vice versa? 'finer than' is not used a lot in this paper. If the two concepts are similar, why not pick one of them?

A2. If $G \trianglelefteq H$, then every granule of H can be expressed as an union of some set of granules in G. If $G \preceq H$, then every granule of G is a subset of some granule of H. These are not equivalent. Consider an example, where G_1 denotes 'weekend' and H_1 denotes 'week'. Then, $G_1 \preceq H_1$, but $G_1 \not\trianglelefteq H_1$. Consider another example, with G_2 denotes 'days' and H_2 denotes 'business-week'. Then, $G_2 \not\preceq H_2$, but $G_2 \not\trianglelefteq H_2$, since each business-week can be

expressed as an union of some days, but Saturdays and Sundays are not subset of any business-week. Further, consider an example with H_3 denoting 'public holidays', then $G \not\preceq H_3$ and $G \not\trianglelefteq H_3$.

Q3. Page 16, Line 13, multiple order-up quasi-circular granularities. The equations for this operation are not trivial to be ignored. Should be added if possible.

A3.

Q4. Can the time granularity concept be extended to continuous time domain?

A4. Linear granularities are defined for continuous time domain. For continuous time domains, the dependent variable is specified for every value of time t, whereas, in applications we generally have the values of the dependent variable for discrete time intervals. Hence, we will restrict ourselves to explore only discrete time domain.

1.4 Other minor issues.

done.