Errata & Addenda

Object Oriented Data Analysis

By J. S. Marron and Ian L. Dryden

Page 52, line 9: Twin Arches 🡪 Tilted Parabolas

Page 64, line 6: from the 🡪 from

Page 97, line -10: shown in Figures 🡪 shown using two different views of the same data set in Figures

Page 118, line 26: positions, and 🡪 positions (nor shapes of the circles of points)

Page 121, line -12: because they represent 🡪 because such directions represent

Page 121, line -11: change, called 🡪 change. The subspace generated by these directions is called

Page 123, line -3: vector-valued data 🡪 vector valued auxiliary data

Page 125, line 4: challenging. In 🡪 challenging and even controversial. In

Page 126, line 20: useful. 🡪 useful, in particular this way of measuring error is the foundation of least squares and the Analysis of Variance. Furthermore this best corresponds to our usual intuitive notion of distance in the three dimensional world in which we live.

Page 126, line -10: The cosine 🡪 On the cosine

Page 126, line -9: other. When 🡪 other. However, it is a proper metric when the data objects lie on the sphere . When

Page 129, line 3: the middle of the 🡪 the

Page 130, Caption of Figure 7.4: invariance. 🡪 invariance of notions of multivariate median.

Page 134, line -16: e.g. as 🡪 e.g. as done through display of curve modes of variation shown in many places starting with Figure 1.4 for the Spanish Mortality data and as

Page 173, Caption of Figure 8.14: indicate fibers 🡪 indicate approximately parallel fibers

Page 180, line 19: MDS using 🡪 MDS (recall Multi Dimensional Scaling from Section 7.2) using

Page 185, Caption of Figure 9.7: Data Objects in 🡪 Piece-wise linear data object curves in

Page 188, line 11: panel. 🡪 panel of Figure 9.9.

Page 197, bottom: Append a new paragraph:

A related research area is currently called *connectome analysis*, where the goal is understanding connections in the human brain. When that is studied at the population level, the issue of what should be the data objects becomes central. Many approaches are based on bundles of fibers. Campbell et al. (2021) have proposed a particularly innovative approach based on Riemannian metrics as data objects. Variation is then studied in the space of Riemannian metrics.

Page 198, line -2: Append a new paragraph: Note that the concept of *tree lines* first introduced in Wang and Marron (2007) was an early version of *modes of variation* as discussed in Section 3.1.4. Another early version are the *tree curves* proposed by Aydin et al. (2012).

Page 199, line 5: operator. 🡪 operator, which led to the inclusion of non-anatomical variation in the data set.

Page 199, line 23: objects. The 🡪 objects. In particular an elegant and very widely used mathematical frame work, often called *BHV Space*, was developed by Billera et al. (2001). The

Page 202, line 1: and Nye 🡪 and see Weyenberg (2015) and Nye

Page 202, end of Section 10.1.2: Append a new paragraph:

A promising recent approach to the analysis of phylogenetic trees as data objects has been proposed by Garba et al. (2021). This revolutionary idea steps completely away from the classical BHV space, through representing the data object trees as correlation matrices (motivated by a new probability framework). The resulting *Wald Space* has far different, yet intuitively sensible properties. Statistical analysis in Wald Space may ultimately prove to be more tractable than in BHV Space because the more direct manifold data analysis methods as described in Section 8.7 may be applicable An interesting open problem is analysis of the Brain Artery data set using Wald Space ideas.

Page 204, line -3: After the Bubenick (2015) reference, add a new sentence: Persistence Landscapes have been integrated with the phase shift - alignment ideas of Chapter 9, in a more recent and interesting analysis of the brain artery data, by Matuk et al. (2021).

Page 350, line 19: studied those 🡪 studied in those

Page 351, line -10: In , the matrix should be .

Page 372, line -15: Marron, J. S. should be a sixth co-author.