# Appendix

**A** Proof of Properties and Theory

**B** Blood Pressure Data

**C** Comparison of Rectangle Plots when range correlation varies from -1 to 1

**D** Comparison of Cross Plots when range correlation varies from -1 to 1

**E** Comparison of Segment Plots when range correlation varies from -1 to 1

**F** Comparison of Guided Dandelion Plots with guiding polygon

**G** Comparison of the rectangle plots when center correlation is -1 and range correlation varies as -1,-0.9,-0.5,0.5,0.9,1

**H** Comparison of the segment plots when center correlation is -1 and range correlation varies as -1,-0.9,-0.5,0.5,0.9,1

**I** Comparison of the guided dandelion plots when center correlation is -1 and range correlation varies as -1,-0.9,-0.5,0.5,0.9,1

**J** Outlier Examples

# A Proof of Properties and Theory

**Proof of Property 1**

We have $V_c = \frac{X_c - \mu_{X_c}}{\sigma_{X_c}}$, and $W_c = \frac{Y_c - \mu_{Y_c}}{\sigma_{Y_c}}$.

Thus $\rho_{cc} = cor(X_c, Y_c) = E[\frac{(X_c-\mu_{X_c})(Y_c-\mu_{Y_c})}{\sigma_{X_c}\sigma_{Y_c}}] = E(V_c W_c)$,

$E[(V_c - W_c)^2] = E(V_c^2 + W_c^2 - 2V_c W_c) = E(V_c^2) + E(W_c^2) - 2E(V_c W_c) = 2 - 2E(V_c W_c) = 2 - 2\rho_{cc}$.

Thus, the length of the projection on the horizontal axis $|V_c - W_c|$ increases as the center correlation $\rho_{cc}$ decreases and vice versa.

**Proof of Theorem 1**

Since $(X_c, Y_c)^T \sim N(\mu, \Sigma)$, where $\mu = (\mu_{X_c}, \mu_{Y_c})^T$, $\Sigma = \begin{pmatrix} \sigma^2_{X_c} & \rho_{cc}\sigma_{X_c}\sigma_{Y_c} \\ \rho_{cc}\sigma_{X_c}\sigma_{Y_c} & \sigma^2_{Y_c} \end{pmatrix}$

$f_{X_c,Y_c}(x_c, y_c) = \frac{1}{2\pi\sqrt{|\Sigma|}} e^{-\frac{1}{2}(x_c-\mu_{X_c}, y_c-\mu_{Y_c})\Sigma^{-1}\begin{pmatrix} x_c - \mu_{X_c} \\ y_c - \mu_{Y_c}\end{pmatrix}}$

Let $V_c = \frac{X_c - \mu_{X_c}}{\sigma_{X_c}}$, $W_c = \frac{Y_c - \mu_{Y_c}}{\sigma_{Y_c}}$, then $X_c = \mu_{X_c} + \sigma_{X_c} V_c$, $Y_c = \mu_{Y_c} + \sigma_{Y_c} W_c$.

$f_{V_c,W_c}(v_c, w_c) = \frac{1}{2\pi\sqrt{|\Sigma|}} e^{-\frac{1}{2}(\sigma_{X_c} v_c, \sigma_{Y_c} w_c)\Sigma^{-1}\begin{pmatrix} \sigma_{X_c} v_c \\ \sigma_{Y_c} w_c\end{pmatrix}} |J|$

Here $|J| = |\begin{pmatrix} \sigma_{X_c} & 0 \\ 0 & \sigma_{Y_c}\end{pmatrix}| = \sigma_{X_c}\sigma_{Y_c}$

So $f_{V_c,W_c}(v_c, w_c) = \frac{1}{2\pi\sqrt{|\Sigma|}} \sigma_{X_c}\sigma_{Y_c} e^{-\frac{1}{2}(v_c,w_c)\begin{pmatrix} \sigma_{X_c} & 0 \\ 0 & \sigma_{Y_c}\end{pmatrix}\Sigma^{-1}\begin{pmatrix} \sigma_{X_c} & 0 \\ 0 & \sigma_{Y_c}\end{pmatrix}\begin{pmatrix} v_c \\ w_c\end{pmatrix}}$

$= \frac{1}{2\pi\sqrt{|\Sigma'|}} \sigma_{X_c}\sigma_{Y_c} e^{-\frac{1}{2}(v_c,w_c)(\Sigma')^{-1}\begin{pmatrix} v_c \\ w_c\end{pmatrix}}$

where $\Sigma' = \begin{pmatrix} \frac{1}{\sigma_{X_c}} & 0 \\ 0 & \frac{1}{\sigma_{Y_c}}\end{pmatrix} \Sigma \begin{pmatrix} \frac{1}{\sigma_{X_c}} & 0 \\ 0 & \frac{1}{\sigma_{Y_c}}\end{pmatrix}$, $(\Sigma')^{-1} = \begin{pmatrix} \sigma_{X_c} & 0 \\ 0 & \sigma_{Y_c}\end{pmatrix} \Sigma \begin{pmatrix} \sigma_{X_c} & 0 \\ 0 & \sigma_{Y_c}\end{pmatrix}$ and $|\Sigma'| = \frac{1}{\sigma^2_{X_c}\sigma^2_{Y_c}}|\Sigma|$

Hence $(V_c, W_c) \sim N\left(\binom{0}{0}, \Sigma' = \begin{pmatrix} 1 & \rho_{cc} \\ \rho_{cc} & 1\end{pmatrix}\right)$.

We have shown that $(V_c, W_c) \sim N\left(\binom{0}{0}, \Sigma' = \begin{pmatrix} 1 & \rho_{cc} \\ \rho_{cc} & 1\end{pmatrix}\right)$.

When $W_c > 0, V_c < 0$ or $V_c > 0, W_c < 0$, the segment intersects the horizontal axis. A segment will either intersect the horizontal axis or not with the probability of intersecting $P(W_c > 0, V_c < 0) + P(V_c > 0, W_c < 0)$, which is a Bernoulli distribution. Hence the number of segments that intersect the horizontal axis follows Binomial distribution with $P(W_c > 0, V_c < 0) + P(V_c > 0, W_c < 0)$.

Let $Z_1, Z_2$ be two independent standard normal random variables.
By Cholesky Decomposition, let $X_1 = Z_1$, $X_2 = \rho Z_1 + \sqrt{1-\rho^2} Z_2$,
then $\binom{X_1}{X_2} \sim N(\binom{0}{0}, \Sigma)$, where $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ We know that $P(X_1 > 0, X_2 < 0) = P(X_1 < 0, X_2 > 0)$.

$$
\begin{aligned}
P(X_1 > 0, X_2 < 0) &= P(Z_1 > 0, Z_2 < \frac{-\rho}{\sqrt{1-\rho^2}} Z_1) \\
&= \int_0^\infty \int_{-\infty}^{\frac{-\rho z_1}{\sqrt{1-\rho^2}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z_1^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z_2^2}{2}} dz_2 dz_1 \\
&= \int_0^\infty \int_{-\infty}^{\frac{-\rho z_1}{\sqrt{1-\rho^2}}} \frac{1}{2\pi} e^{-\frac{1}{2}(z_1^2+z_2^2)} dz_1 dz_2
\end{aligned}
$$

Let $Z_1 = rcos(\theta)$ and $Z_2 = rsin\theta$. Then $Z_1^2 + Z_2^2 = r^2$ and $\frac{Z_2}{Z_1} = tan\theta$.
By switching to the polar system, we have the following.

$$
\begin{aligned}
P(X_1 > 0, X_2 < 0) &= P(Z_1 > 0, Z_2 < \frac{-\rho}{\sqrt{1-\rho^2}} Z_1) \\
&= \int_{-\frac{\pi}{2}}^{tan^{-1}(\frac{-\rho}{\sqrt{1-\rho^2}})} \int_0^\infty \frac{1}{2\pi} e^{-\frac{1}{2}(r^2)} r dr d\theta \\
&= \int_{-\frac{\pi}{2}}^{tan^{-1}(\frac{-\rho}{\sqrt{1-\rho^2}})} \frac{1}{2\pi} d\theta \\
&= \frac{1}{2\pi}(tan^{-1}(\frac{-\rho}{\sqrt{1-\rho^2}}) + \frac{\pi}{2}) \\
&= \frac{1}{4} - \frac{1}{2\pi} tan^{-1}(\frac{\rho}{\sqrt{1-\rho^2}})
\end{aligned}
$$

Let $\phi \in [-\pi/2, \pi/2]$, and $\phi = tan^{-1}(\frac{\rho}{\sqrt{1-\rho^2}})$. Then $tan(\phi) = \frac{\rho}{\sqrt{1-\rho^2}}$ and $sin(\phi) = \rho$ and $\phi = sin^{-1}(\rho)$

Hence

$$
\begin{aligned}
P(X_1 > 0, X_2 < 0) &= \frac{1}{4} - \frac{1}{2\pi} tan^{-1}(\frac{\rho}{\sqrt{1-\rho^2}}) \\
&= \frac{1}{4} - \frac{1}{2\pi} sin^{-1}(\rho)
\end{aligned}
$$

So,

$$
\begin{aligned}
P(X_1 > 0, X_2 < 0) + P(X_1 < 0, X_2 > 0) &= \frac{1}{2} - \frac{1}{\pi} sin^{-1}(\rho) \\
&= \frac{1}{2} - \frac{1}{\pi}(\frac{\pi}{2} - cos^{-1}(\rho)) \\
&= cos^{-1}(\rho)/\pi
\end{aligned}
$$

Hence we have $E(Z_c) = n * (P(V_c < 0, W_c > 0) + P(V_c > 0, W_c < 0)) = n * \frac{cos^{-1}\rho_{cc}}{\pi}$.

**Proof related to Folded Normal Distribution**

We know that $V_c, W_c, V_r$ and $W_r$ follow standard normal distribution independently.

Hence $E(V_c - W_c) = E(V_c) - E(W_c) = 0$ and $Var(V_c - W_c) = Var(V_c) + Var(W_c) = 2$. And since $V_c - W_c$ is a linear combination of two independent normal distribution, $V_c - W_c \sim N(0, 2)$.

Similarly, we have $V_c - W_r + V_r - W_c \sim N(0, 4)$ and $V_c - W_r - V_r + W_c \sim N(0, 4)$.

So $V_c - W_c$, $V_c - W_r + V_r - W_c$ and $V_c - W_r - V_r + W_c$ all follow folded normal distribution.

Let X be a random variable follow normal distribution with mean $\mu$ and variance $\sigma^2$. Then $Y = |X|$ follow folded normal distribution.

Let Z be a standard normal distribution, then $Y = |\mu + \sigma Z|$.

So

$$
\begin{aligned}
E(Y) &= E(|\mu+\sigma Z|) \\
&= E(\mu+\sigma Z; \mu+\sigma Z \geq 0) - E(\mu+\sigma Z; \mu+\sigma Z < 0) \\
&= E(\mu+\sigma Z; \mu+\sigma Z \geq 0) + E(\mu+\sigma Z; \mu+\sigma Z < 0) - 2E(\mu+\sigma Z; \mu+\sigma Z < 0) \\
&= E(\mu+\sigma Z) - 2E(\mu+\sigma Z; Z < -\mu/\sigma) \\
&= \mu - 2E(\mu; Z < -\mu/\sigma) - 2E(\sigma Z; Z < -\mu/\sigma) \\
&= \mu - 2\mu\Phi(-\mu/\sigma) - 2\sigma E(Z; Z < -\mu/\sigma)
\end{aligned}
$$

Here

$$
E(Z; Z < -\mu/\sigma) = \int_{-\infty}^{-\mu/\sigma} z \frac{1}{\sqrt{2\pi} e^{-z^2/2}} dz
$$

$$
\text{Let } u = z^2/2 \text{ and } d_u = z dz
$$

$$
\begin{aligned}
&= \int_{\infty}^{\mu^2/2\sigma^2} \frac{1}{\sqrt{2\pi}} e^{-u} du \\
&= -\frac{1}{\sqrt{2\pi}} e^{-u} \Big|_{\infty}^{\mu^2/2\sigma^2} \\
&= -\frac{1}{\sqrt{2\pi}} e^{-\frac{\mu^2}{2\sigma^2}}
\end{aligned}
$$

So $E(Y) = \mu - 2\mu\Phi(-\mu/\sigma) + 2\frac{1}{\sqrt{2\pi}} e^{-\frac{\mu^2}{2\sigma^2}}$

When $\mu = 0$ and $\sigma^2 = 2$, $E(|V_c - W_c|) = \frac{2}{\sqrt{\pi}}$.

And when $\mu = 0$ and $\sigma^2 = 4$, $E(|V_c - W_r + V_r - W_c|) = \frac{2\sqrt{2}}{\sqrt{\pi}}$, $E(|V_c - W_r - V_r + W_c|) = \frac{2\sqrt{2}}{\sqrt{\pi}}$.

Also

$$
E(Y^2) = E(|X|^2) = E(X^2) = Var(X) + (E(X))^2 = \sigma^2 + \mu^2
$$

So

$$
Var(Y) = E(Y^2) - (E(Y))^2 = \sigma^2 + \mu^2 - (\mu - 2\mu\Phi(-\mu/\sigma) + 2\frac{1}{\sqrt{2\pi}} e^{-\frac{\mu^2}{2\sigma^2}})^2
$$

When $\mu = 0$ and $\sigma^2 = 2$, $Var(|V_c - W_c|) = 2 - \frac{4}{\pi}$.

# B Blood Pressure Data

| | Pulse Rate | Systolic Pressure |
|---|---|---|
| 1 | [44,68] | [90,100] |
| 2 | [60,72] | [90,130] |
| 3 | [56,90] | [140,180] |
| 4 | [70,112] | [110,142] |
| 5 | [54,72] | [90,100] |
| 6 | [70,100] | [130,160] |
| 7 | [72,100] | [130,160] |
| 8 | [76,98] | [110,190] |
| 9 | [86,96] | [138,180] |
| 10 | [86,100] | [110,150] |
| 11 | [63,75] | [60,100] |

Table B.1: Blood pressure data of 11 patients. Here each column is in interval format. (Billard and Diday, 2000)

# C Comparison of Rectangle Plots (when range correlation varies from -1 to 1)

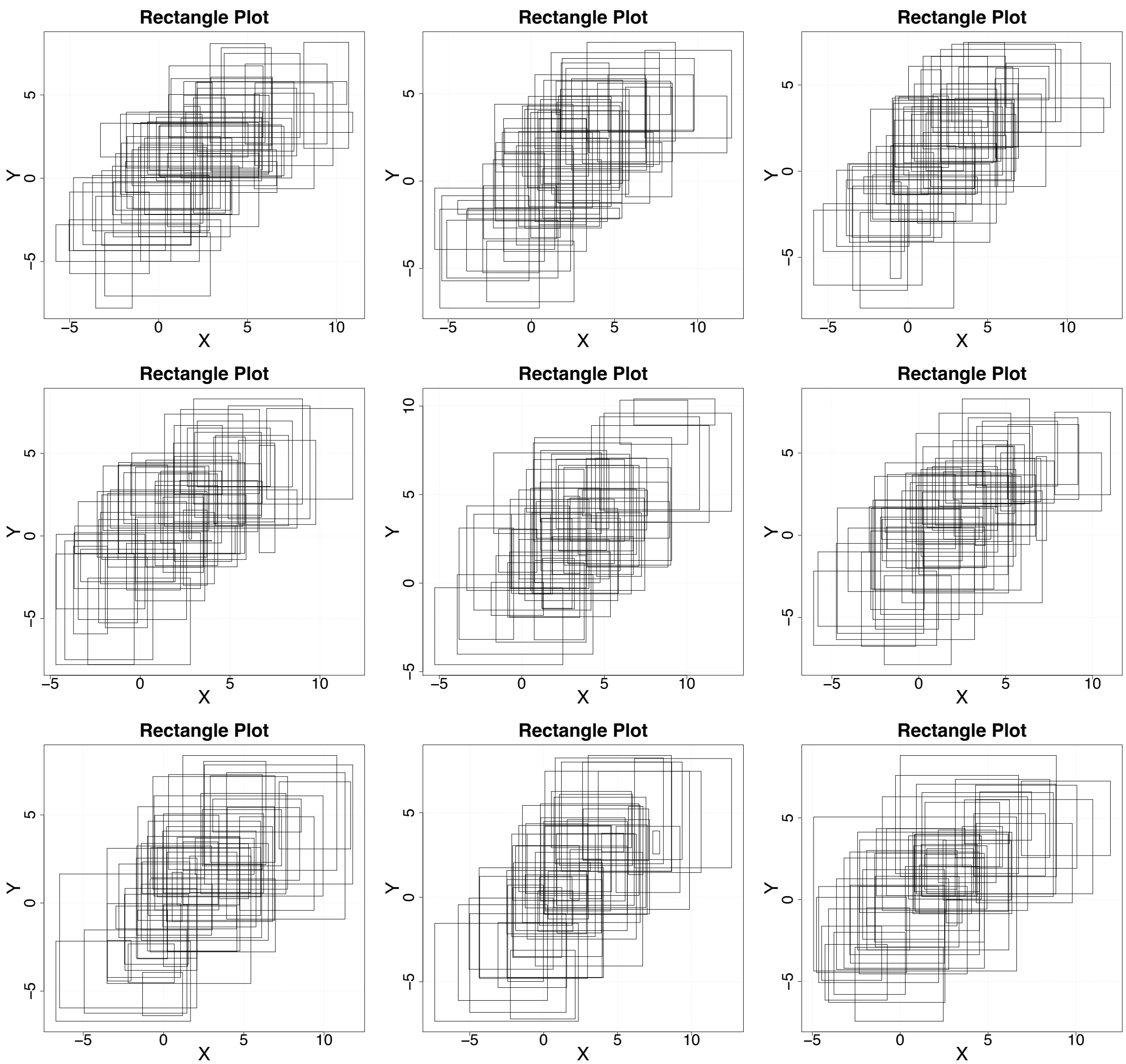


Figure C.1: Rectangle plot for 9 different range correlations change from -1 to 1 with the exact same center correlation $\rho_{cc} = 0.8$.

# D Comparison of Cross Plots (when range correlation varies from -1 to 1)

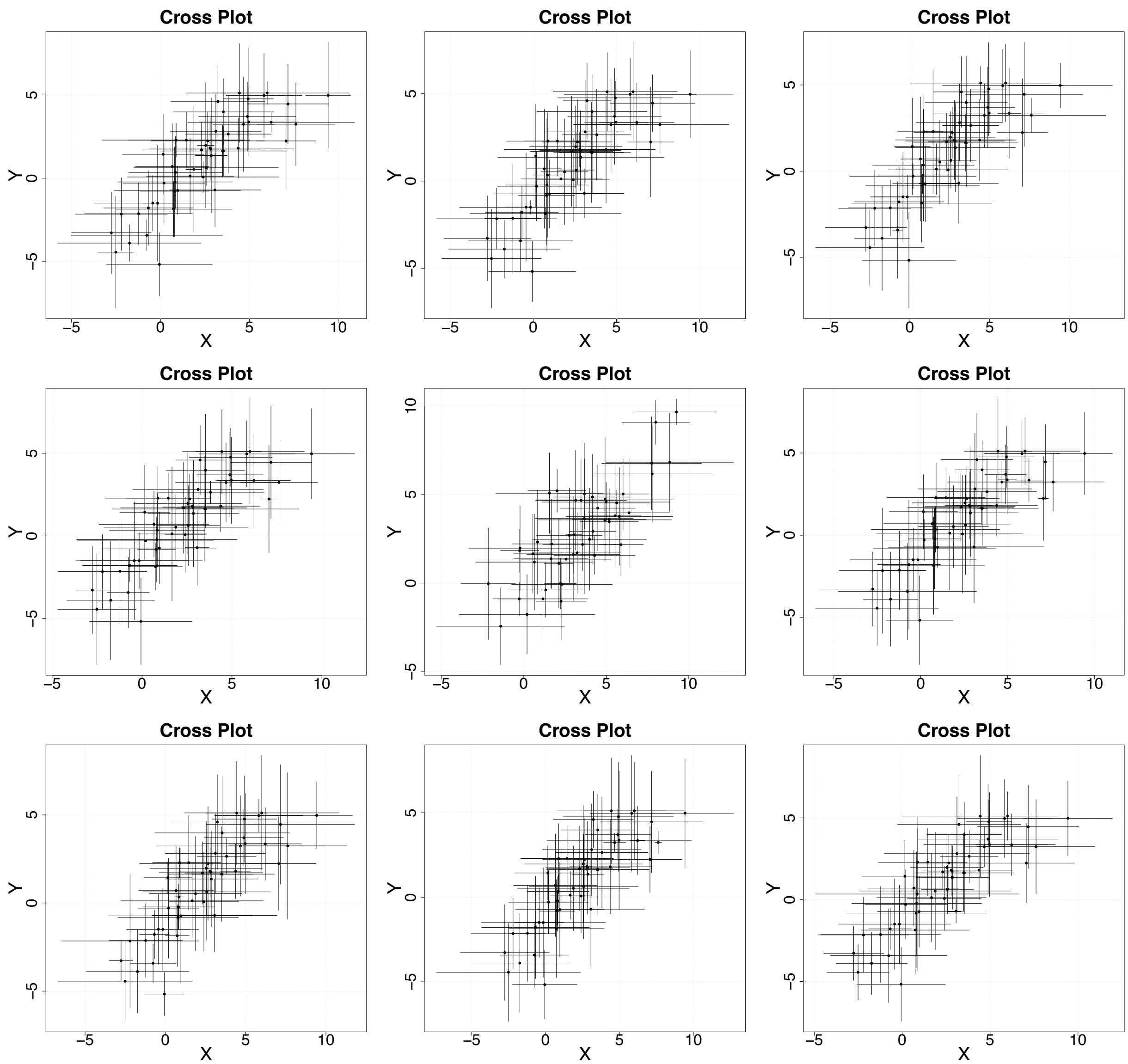


Figure D.1: Cross plots for the case where center correlations remain as 0.8 and range correlations change from -1 to 1

# E Comparison of Segment Plots (when range correlation varies from -1 to 1)

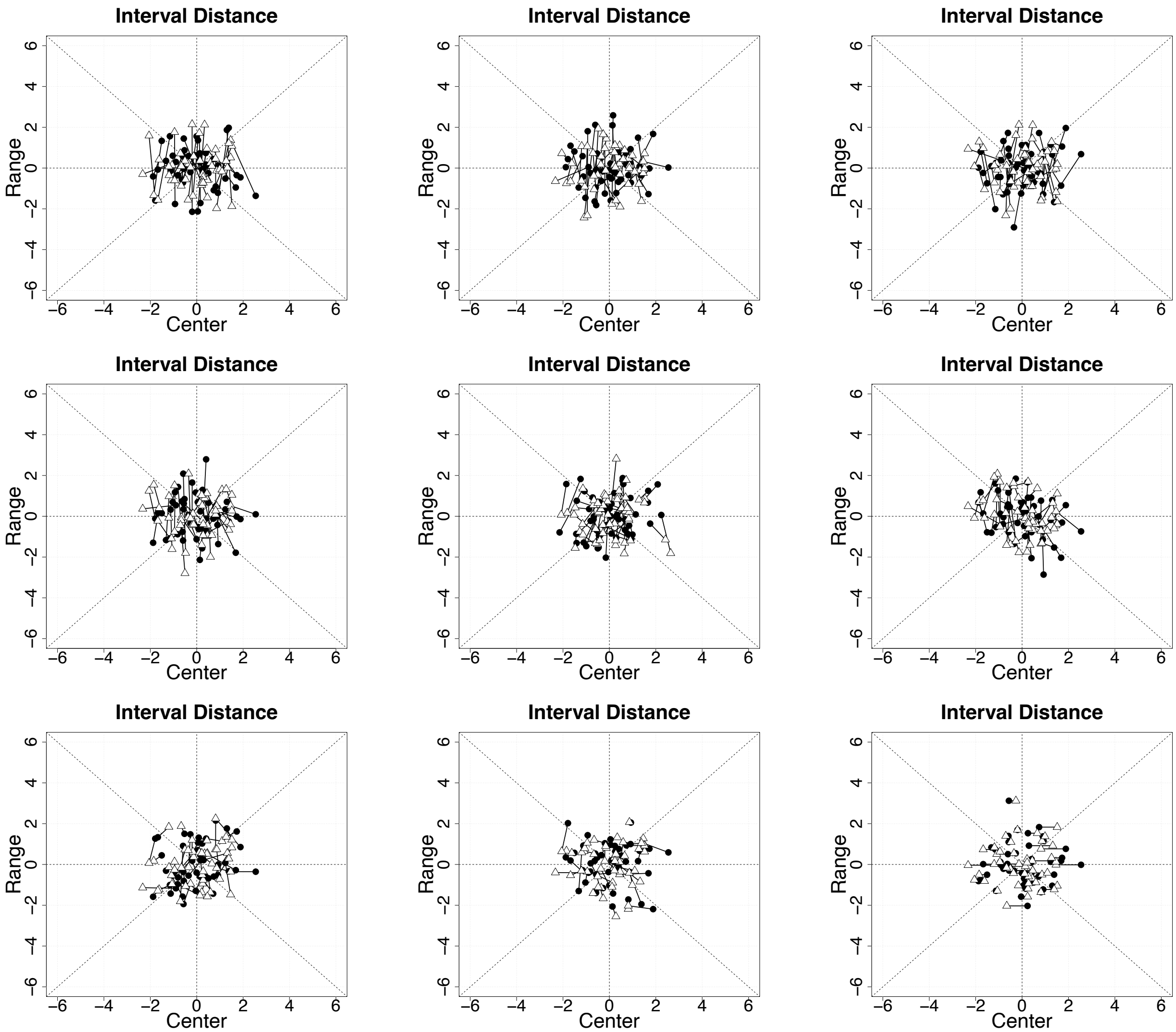


Figure E.1: Segment plots for 9 different range correlations change from -1 to 1 with the exact same center correlation $\rho_{cc} = 0.8$. The solid circle points correspond to the variable $X$; and the hollow triangle points correspond to the variable $Y$.

# F Comparison of Guided Dandelion Plots with guiding polygon (when range correlation varies from -1 to 1)

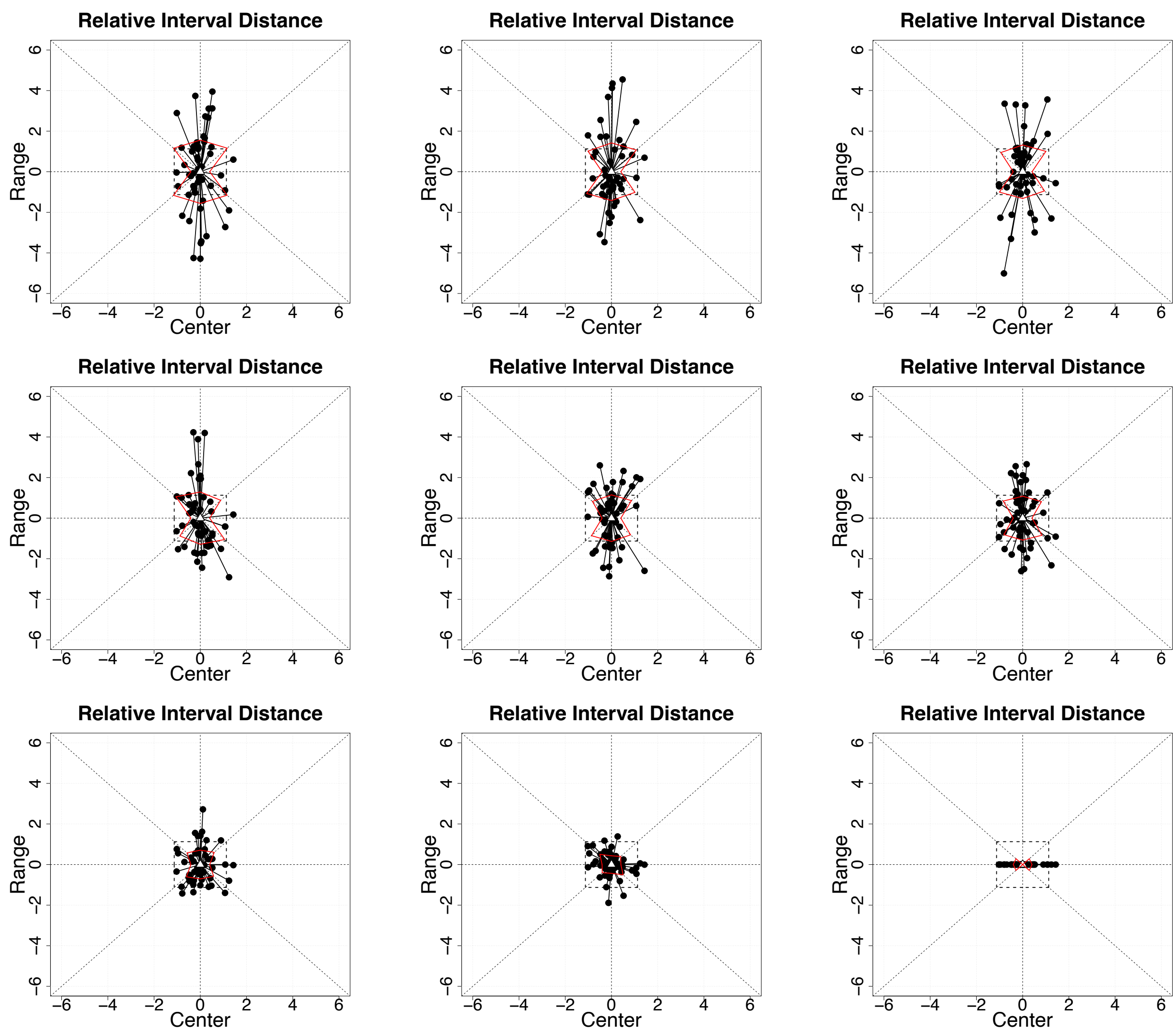


Figure F.1: Guided dandelion plot for 9 different range correlations change from -1 to 1 with the exact same center correlation $\rho_{cc} = 0.8$. The solid circle points correspond to the variable $X$; and the hollow triangle points correspond to the variable $Y$.

# G Comparison of the rectangle plots (when center correlation is -1 and range correlation varies as -1,-0.9,-0.5,0.5,0.9,1)

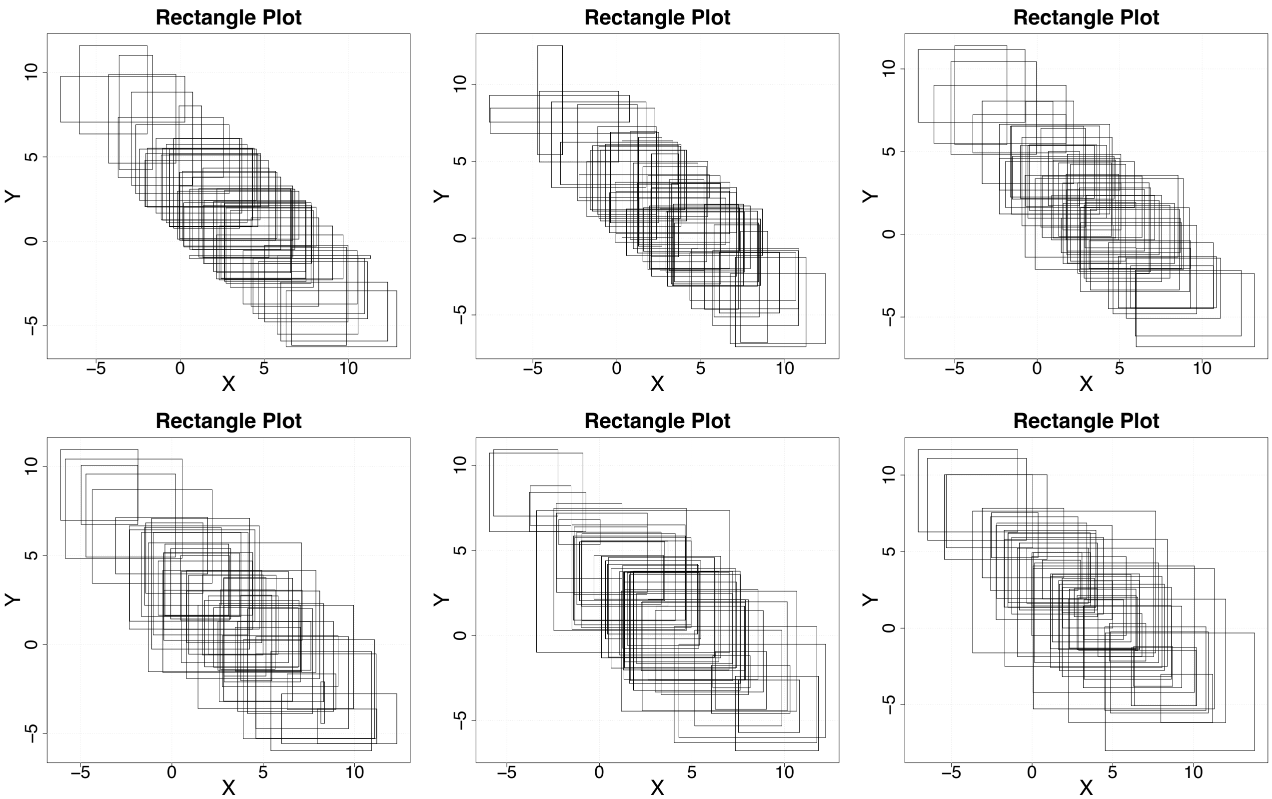


Figure G.1: Rectangle plots with $\rho_{cc} = -1$ and $\rho_{rr} = -1, -0.9, -0.5, 0.5, 0.9, 1$

## H Comparison of the segment plots (when center correlation is -1 and range correlation varies as -1,-0.9,-0.5,0.5,0.9,1)

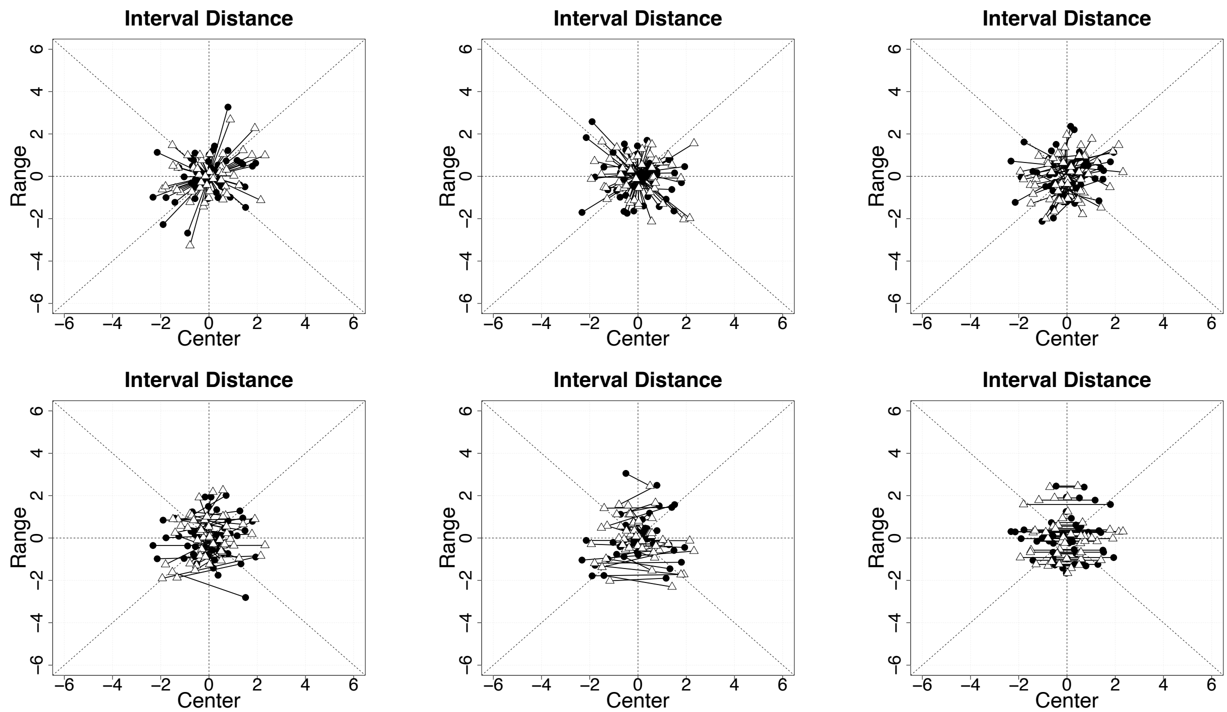


Figure H.1: Segment plots with $\rho_{cc} = -1$ and $\rho_{rr} = -1, -0.9, -0.5, 0.5, 0.9, 1$. The solid circle points correspond to the variable $X$; and the hollow triangle points correspond to the variable $Y$.

# I Comparison of the guided dandelion plots (when center correlation is -1 and range correlation varies as -1,-0.9,-0.5,0.5,0.9,1)

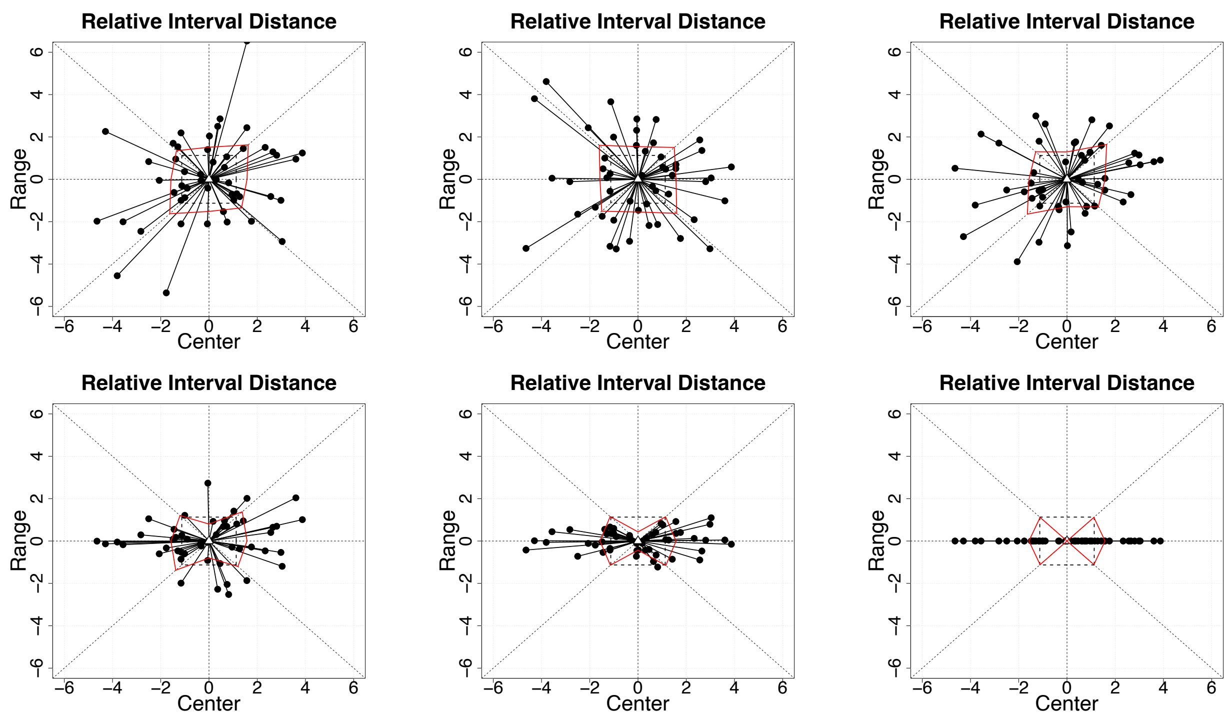


Figure I.1: Guided Dandelion plots with $\rho_{cc} = -1$ and $\rho_{rr} = -1, -0.9, -0.5, 0.5, 0.9, 1$. The solid circle points correspond to the variable $X$; and the hollow triangle points correspond to the variable $Y$.

# J Outlier Example

Here we provide two examples for outlier detection. Figure J.1 displays a rectangle plot, a segment plot and a dandelion plot for a dataset with one extreme value in $X_r$. In the segment plot, there is a solid circle point that has a longer projection on the vertical axis. This implies that there is an outlier in $X_r$. Meanwhile, it is difficult to detect the extreme value in the rectangle plot.

In the second example, we consider the case where a highly positive linear relationship exists between $X_r$ and $Y_r$ and one pair of intervals departs from the relationship. Figure J.2 displays the rectangle plot, segment plot and dandelion plot. From Figure J.2, the rectangle plot cannot clearly detect such pair of intervals. In the segment plot, because the values of the intervals are not extreme, no obvious outliers can be detected. In the dandelion plot, one point has extremely long length of projection on the vertical axis. This point represents the outlier as circled in the plot.

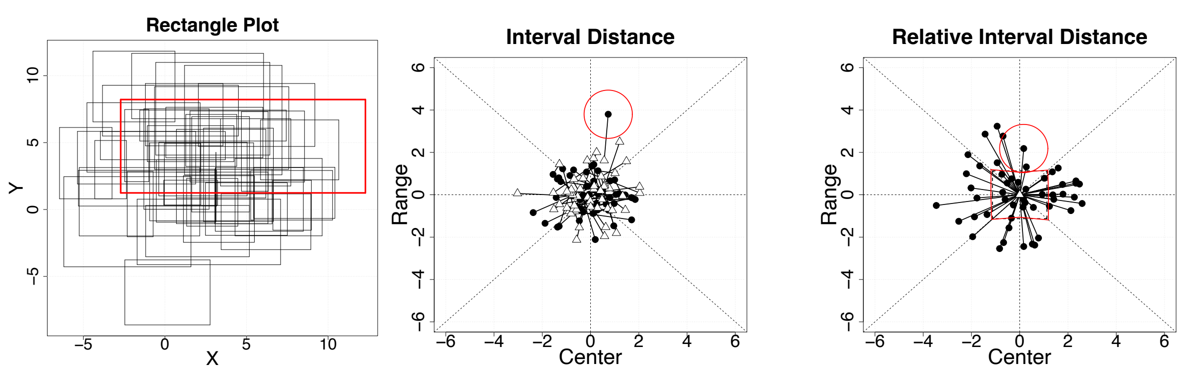


Figure J.1: Scenario where there is an extreme value in $X_r$. The circled point in the segment plot and dandelion plot indicates the outlier.

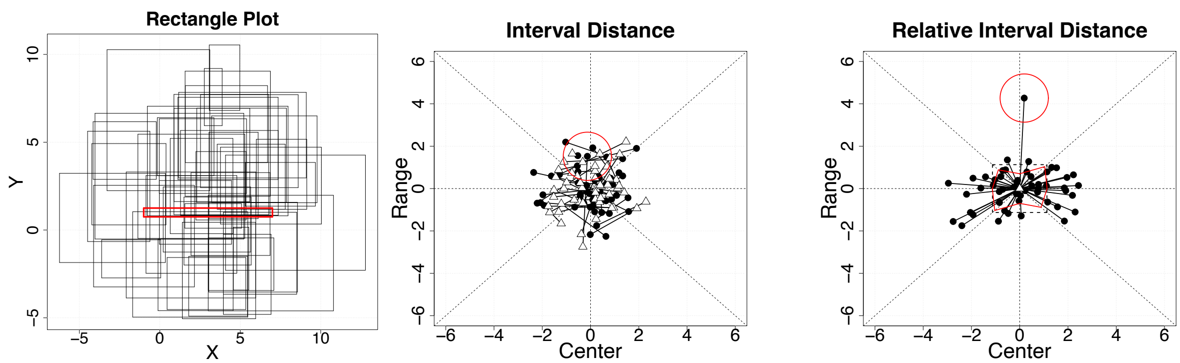


Figure J.2: From the left to the right are the rectangle plot, segment plot and dandelion plot respectively. This is corresponding to the example where $X_r$ and $Y_r$ has 0.8 correlation and there is one pair of intervals departing from the relationship.