1 Covariance

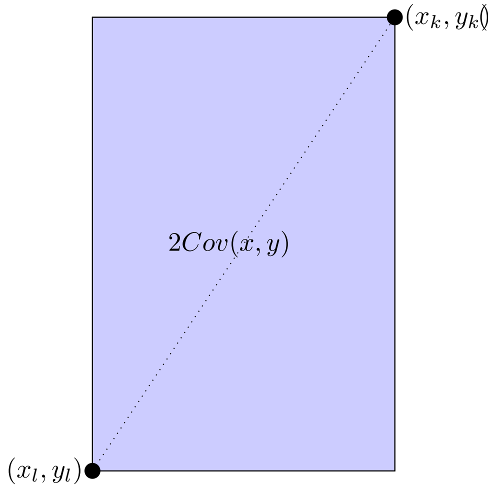


Figure 1: Case 1 : Corrélation positive

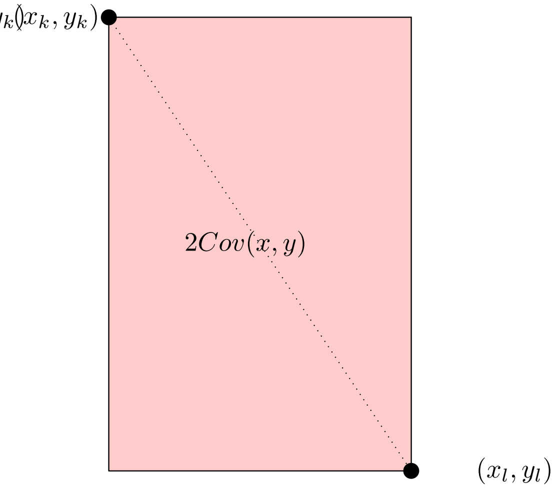
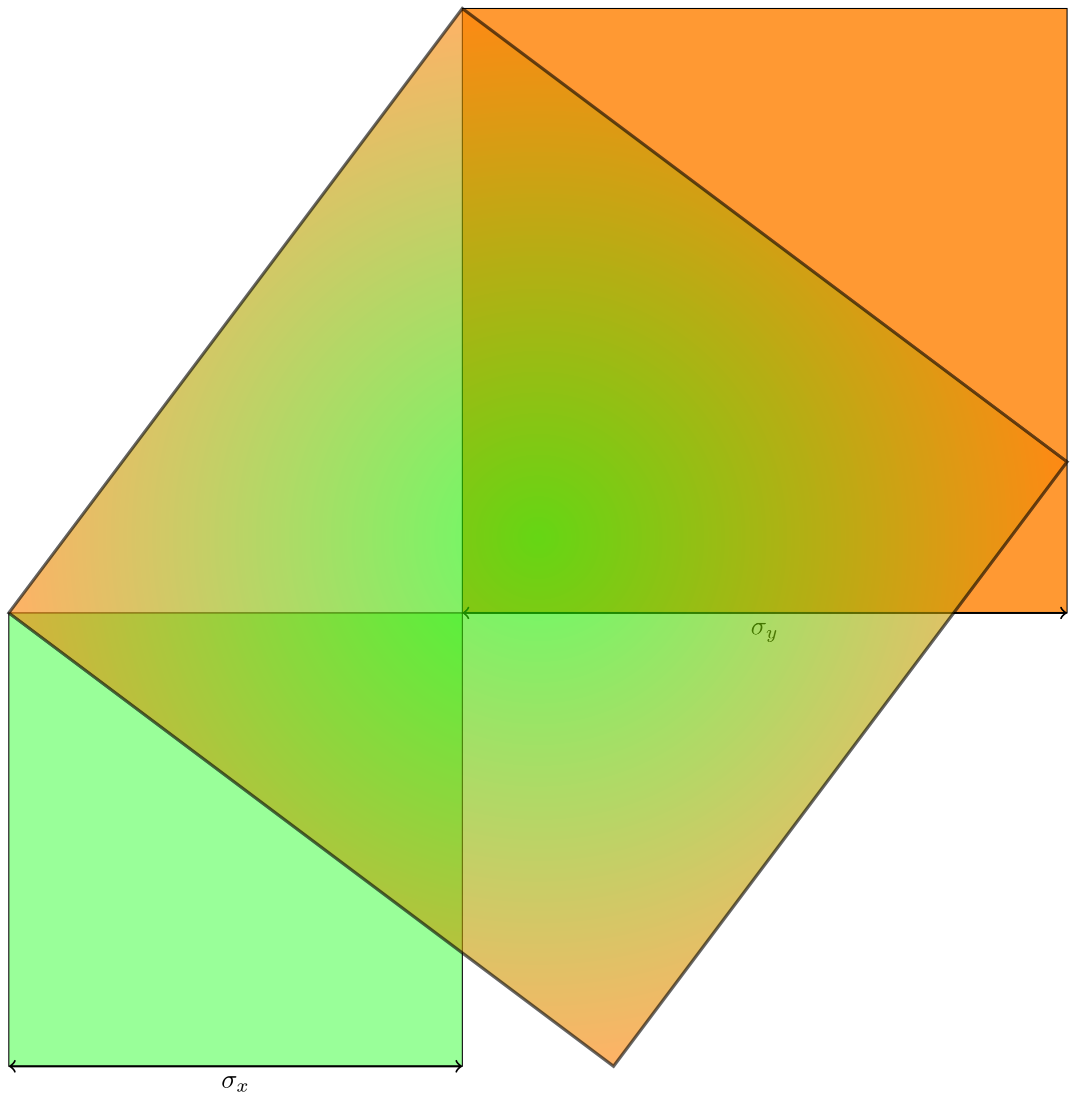
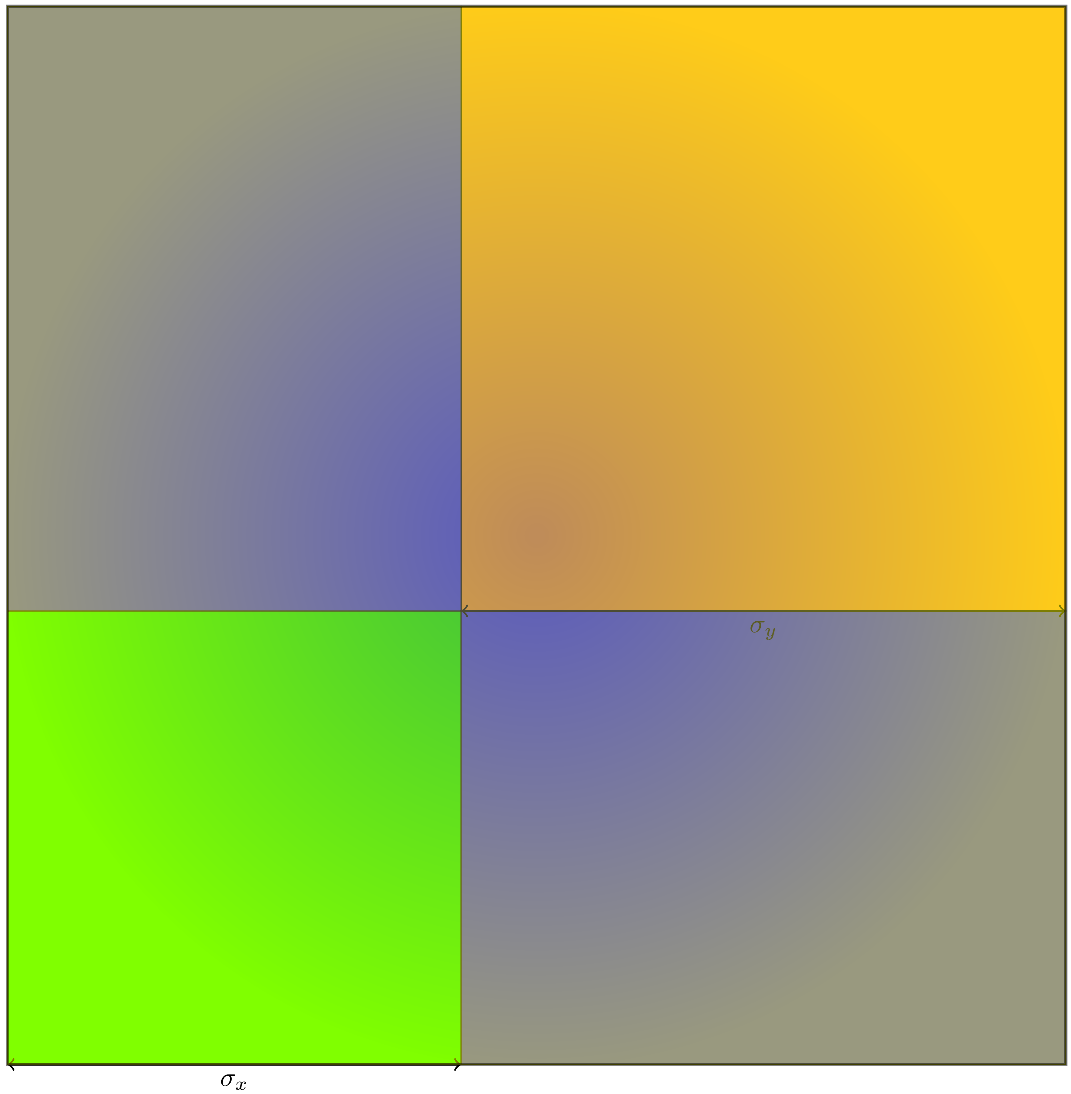


Figure 1: Case 2 : Corrélation négative

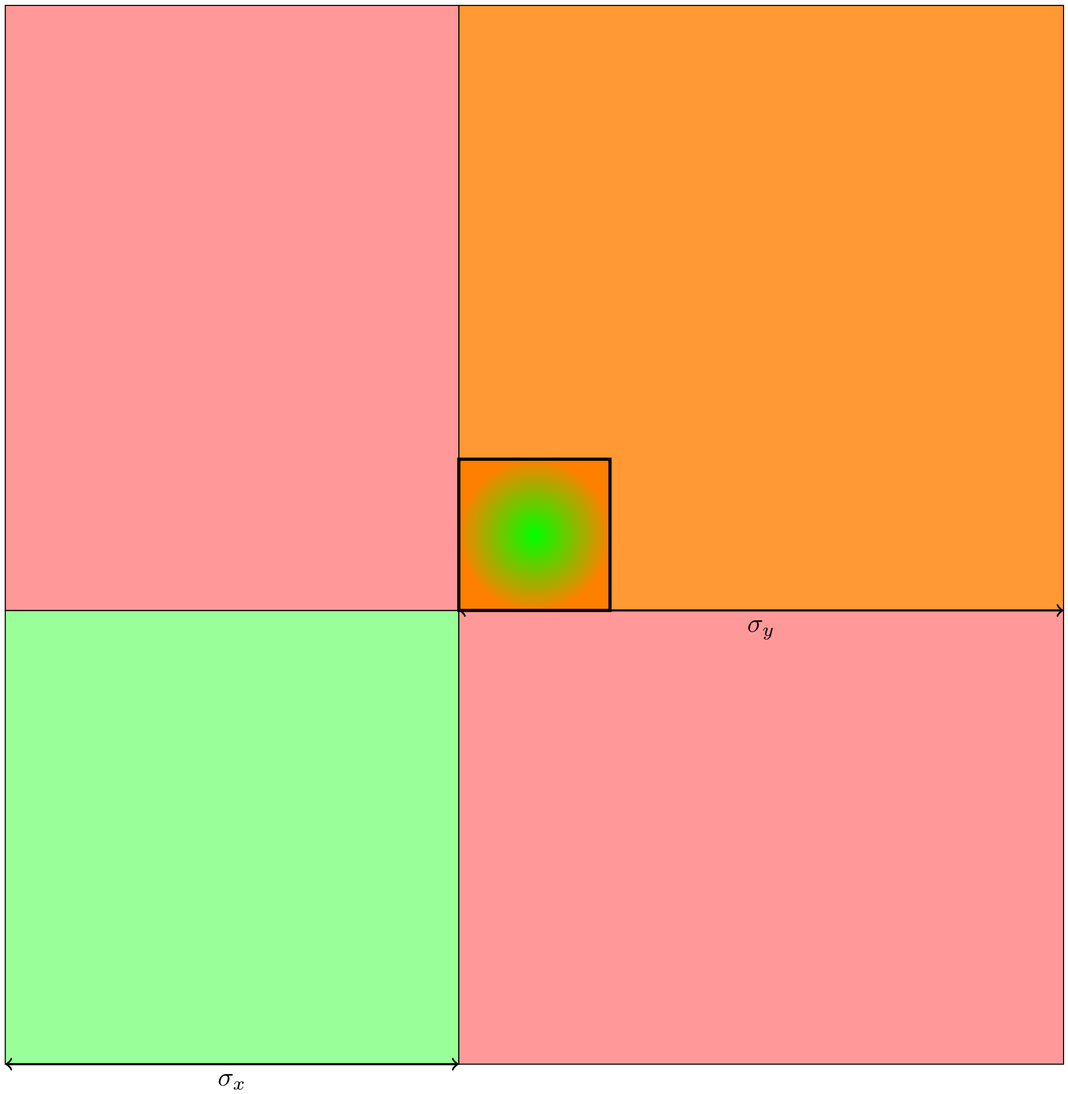
2 $Var(X+Y) = \sigma_x^2 + \sigma_y^2$



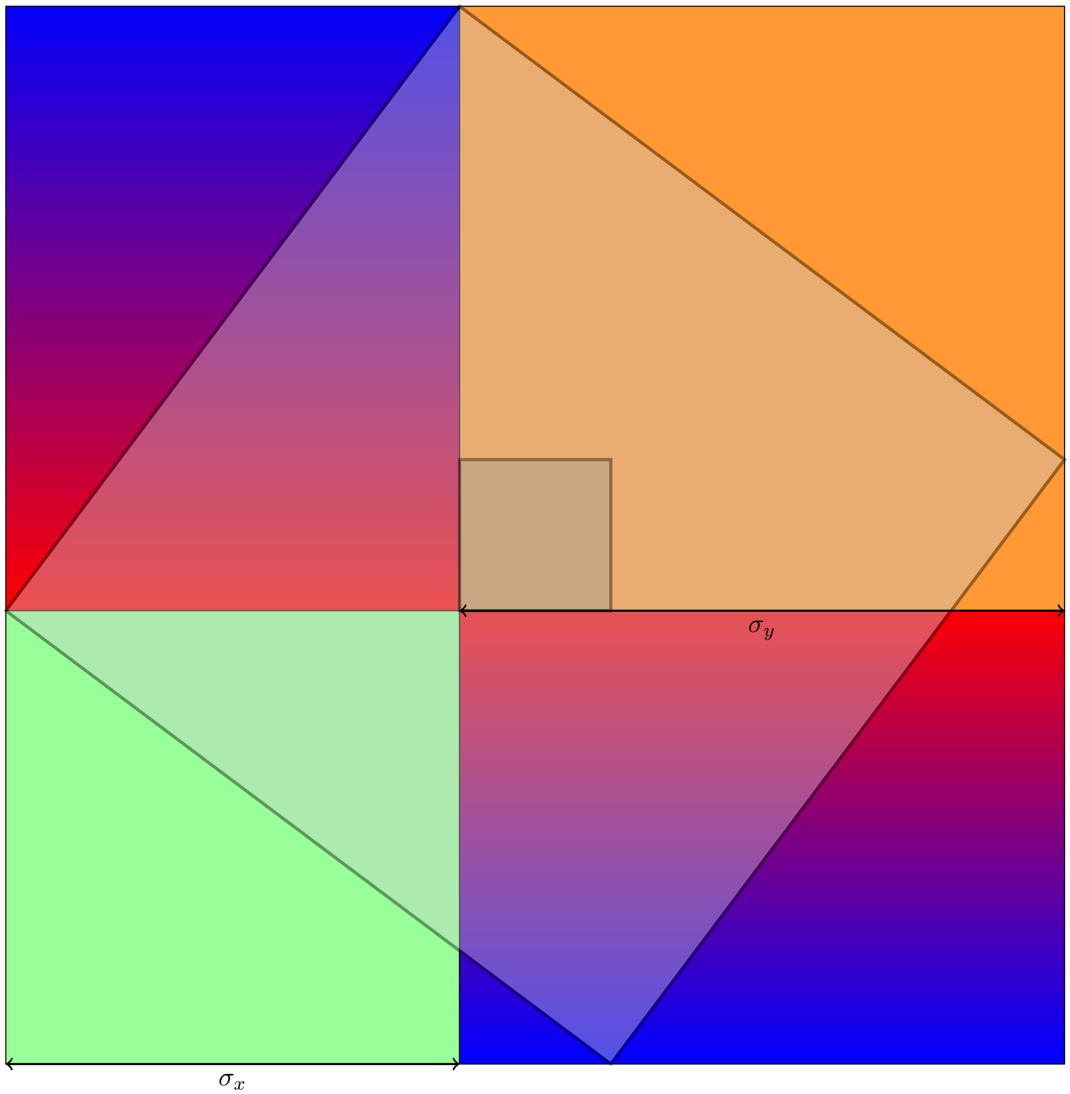
3 $Var(X+Y) = \sigma_x^2 + \sigma_y^2 + 2\sigma_x\sigma_y$



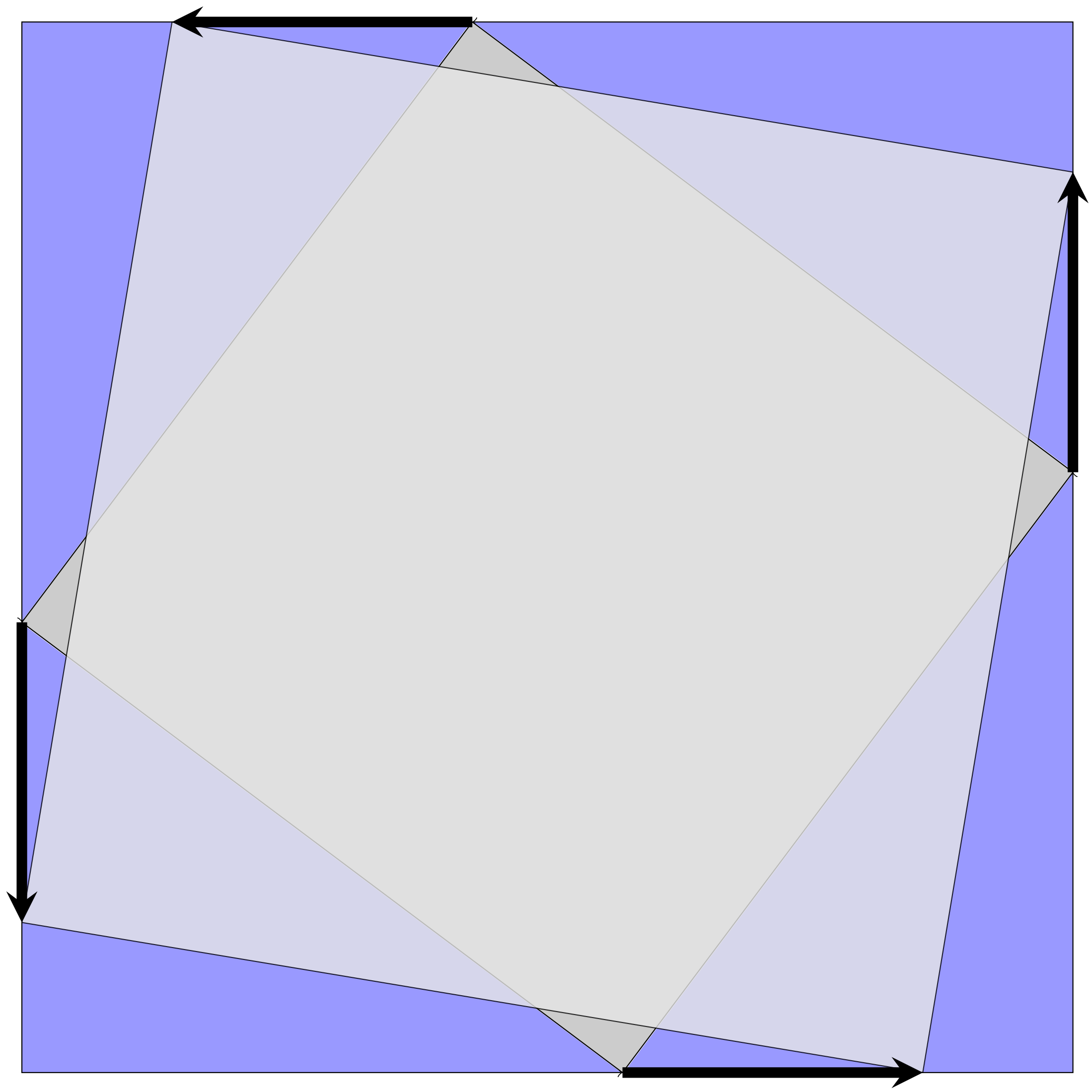
4 $Var(X+Y) = \sigma_x^2 + \sigma_y^2 - 2\sigma_x\sigma_y$



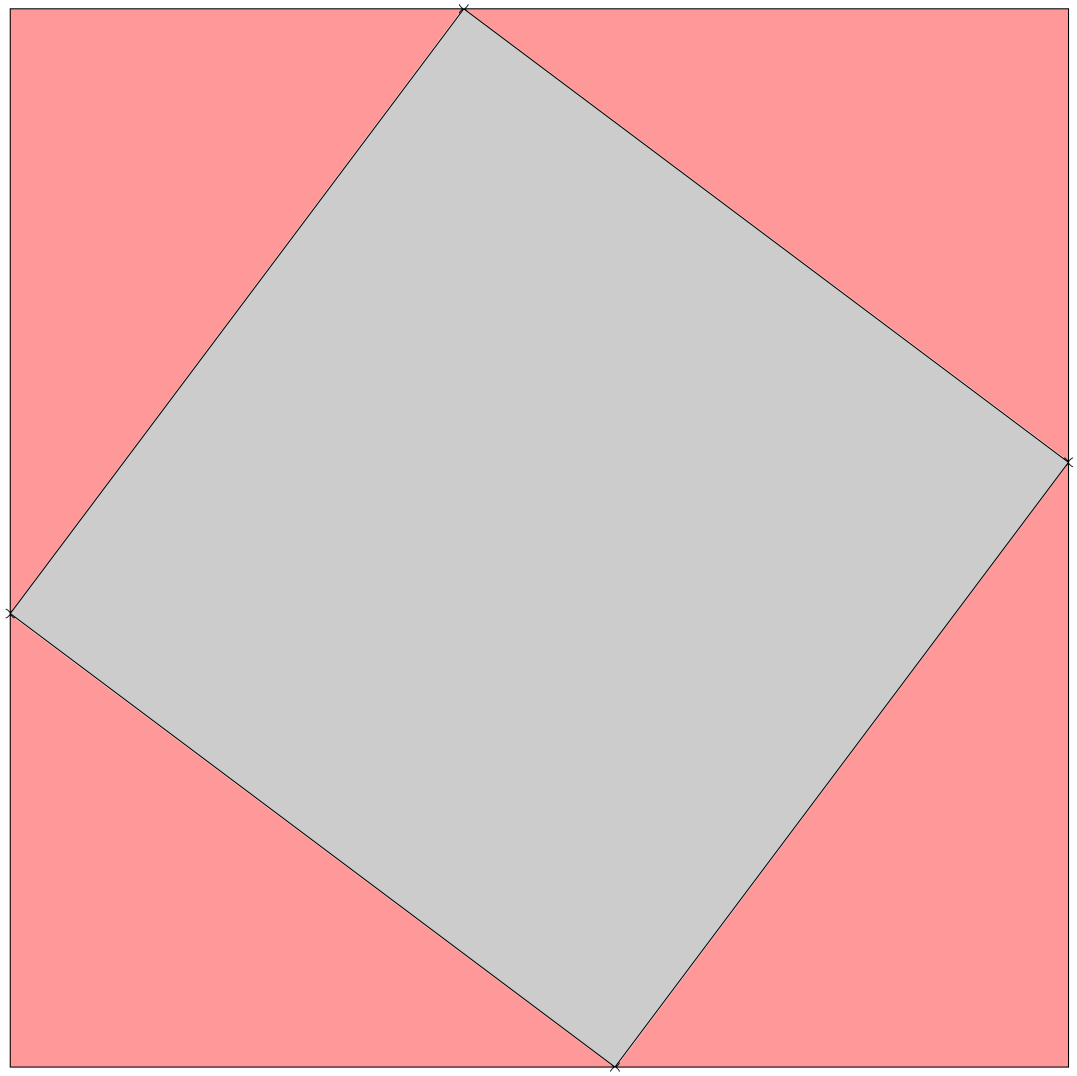
5 $|Cov(X,Y)| \leq \sqrt{\sigma_x^2 + \sigma_y^2}$



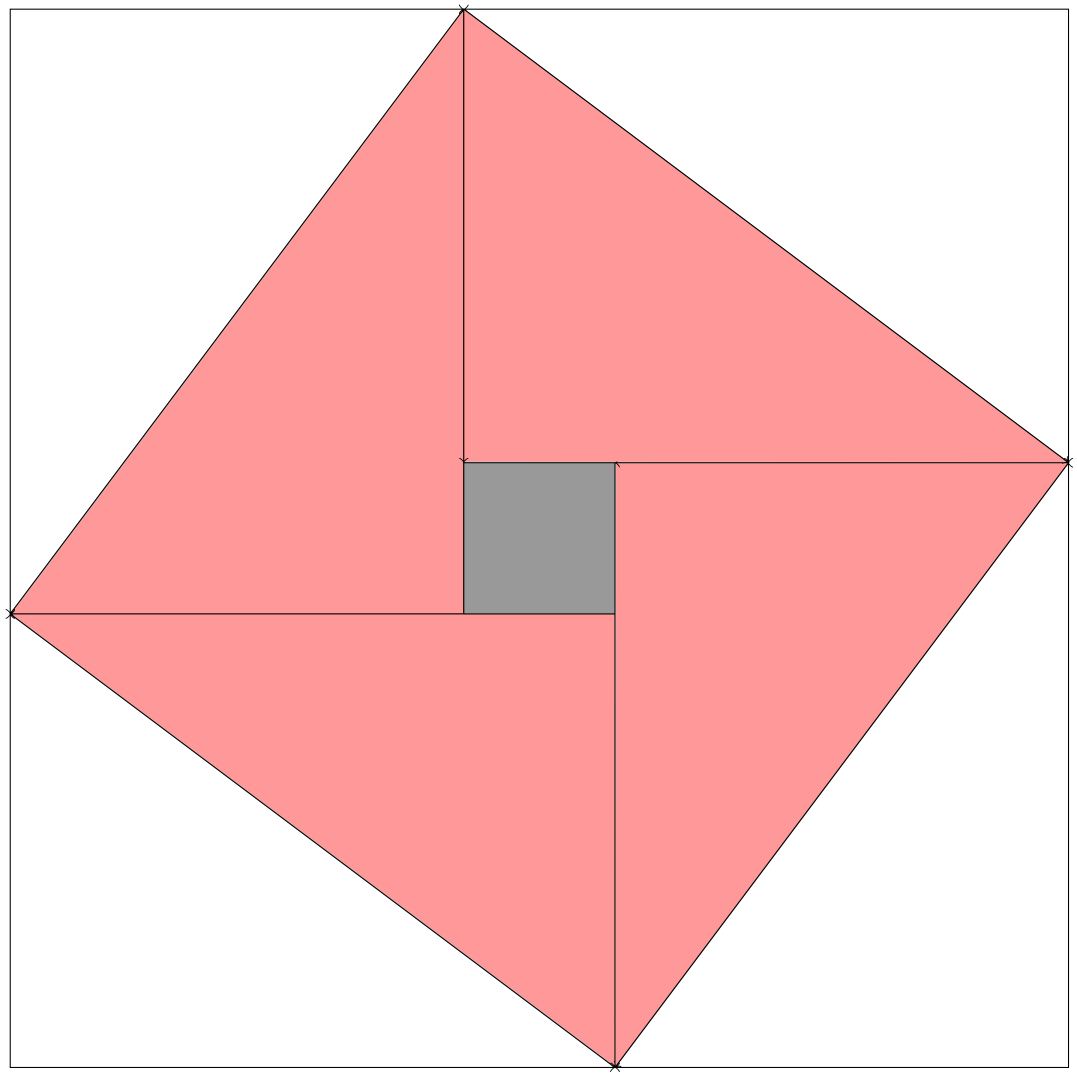
6 Positive covariance



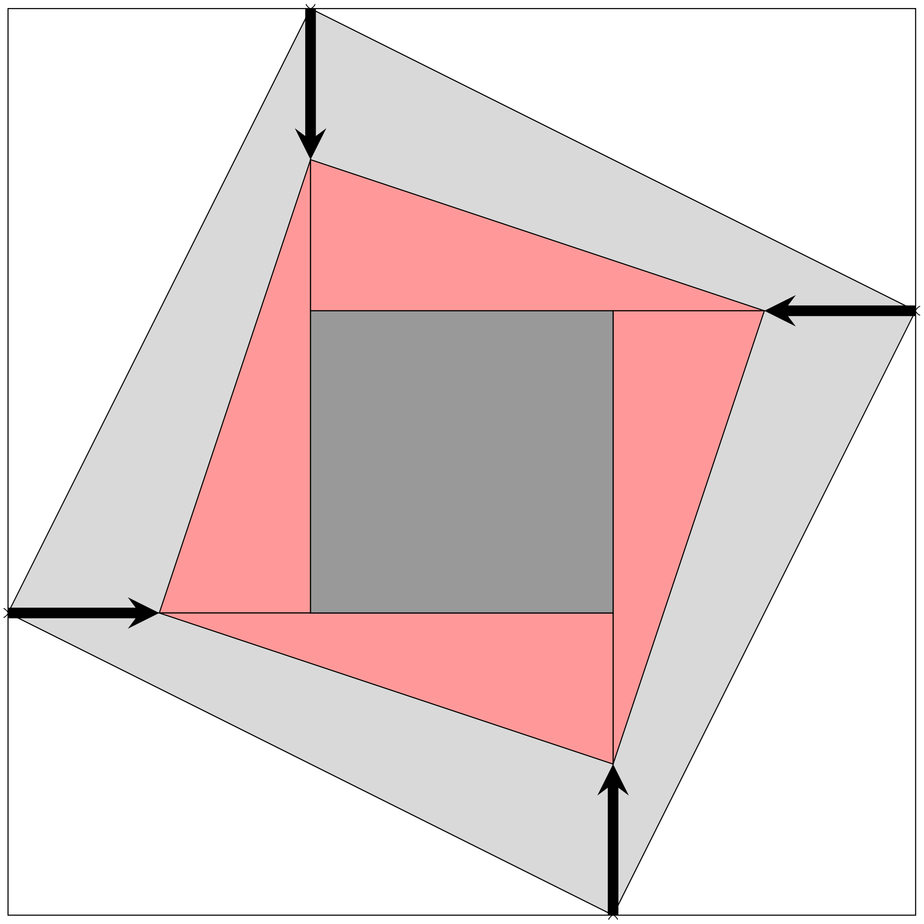
7 Negative covariance first representation...!



8 Negative covariance, perfect negative correlation



9 Negative covariance



10 Un peu de calculs

Variance de la somme de deux variables aléatoires.

$$\begin{aligned} Var(X+Y) &= Cov(X+Y, X+Y) \\ &= E((X+Y)^2) - E(X+Y)E(X+Y) \end{aligned}$$

en développant,

$$\begin{aligned} &= E(X^2) - (E(X))^2 + E(Y^2) - (E(Y))^2 + 2(E(XY) - E(X)E(Y)) \\ &= Var(X) + Var(Y) + 2(E(XY)) - E(X)E(Y)) \\ &= Var(X) + Var(Y) + 2Cov(X,\ Y) \end{aligned}$$

Si les variables $X,\ Y$ sont indépendantes $E(XY) = E(X)E(Y)$ on alors
$Var(X+Y) = Var(X) + Var(Y)$

Limite de $Cov(X,\ Y)$

Soit, $Var(\lambda X + Y)$ où $\lambda \in \mathbb{R}$
$Var(\lambda X + Y) = \lambda^2 Var(X) + 2\lambda Cov(X,\ Y) + Var(Y)$
Comme une variance est positive ou nulle, le polynôme du second degré obtenu $P(\lambda) = Var(\lambda X + Y)$ a un discriminant négatif ou nul.

$$\begin{aligned} \Delta &= 4Cov(X,\ Y)^2 - 4Var(X)Var(Y) \\ &= 4[Cov(X,\ Y)^2 - Var(X)Var(Y)] \leq 0 \end{aligned}$$

Donc $Cov(X,Y)^2 \leq Var(X)Var(Y)$ [1]
ou encore $-\sigma_x\sigma_y \leq Cov(X,Y) \leq \sigma_x\sigma_y$.
Remarquons que si $\Delta = 0$ alors on a l'égalité $Cov(X,Y)^2 = Var(X)Var(Y)$, i.e. s'il existe λ tel que $Var(\lambda X + Y) = 0$
Mais pour conclure, si l'on a $\lambda X + Y$ égal avec probabilité 1 à une constante, disons c, c'est bien que $Y = c - \lambda X$ presque surement, autrement dit un lien linéaire parfait entre les deux variables.

Définition du coefficient de corrélation linéaire (Bravais-Pearson)

- cas $Cov(X,\ Y) \geq 0$

On définit $\rho = \frac{2Cov(X,\ Y)}{Var_{max} - Var_{min}}$

$Var_{max} = \sigma_X^2 + \sigma_Y^2 + 2\sigma_X\sigma_Y$

[1] En théorie des probabilités, l'inégalité de Cauchy-Schwarz permet la preuve du résultat en raison de l'inégalité $E(XY) \leq \sqrt{E(X^2)E(Y^2)}$

$Var_{min} = \sigma_X^2 + \sigma_Y^2$

$Var_{max} - Var_{minx} = 2\sigma_X \sigma_Y$

d'où en simplifiant

$\rho = \frac{Cov(X,\ Y)}{\sigma_X \sigma_Y}$

- cas $Cov(X,\ Y) \leq 0$

On définit $\rho = \frac{2Cov(X,\ Y)}{Var_{max} - Var_{min}}$

$Var_{max} = \sigma_X^2 + \sigma_Y^2$

$Var_{min} = \sigma_X^2 + \sigma_Y^2 - 2\sigma_X \sigma_Y$

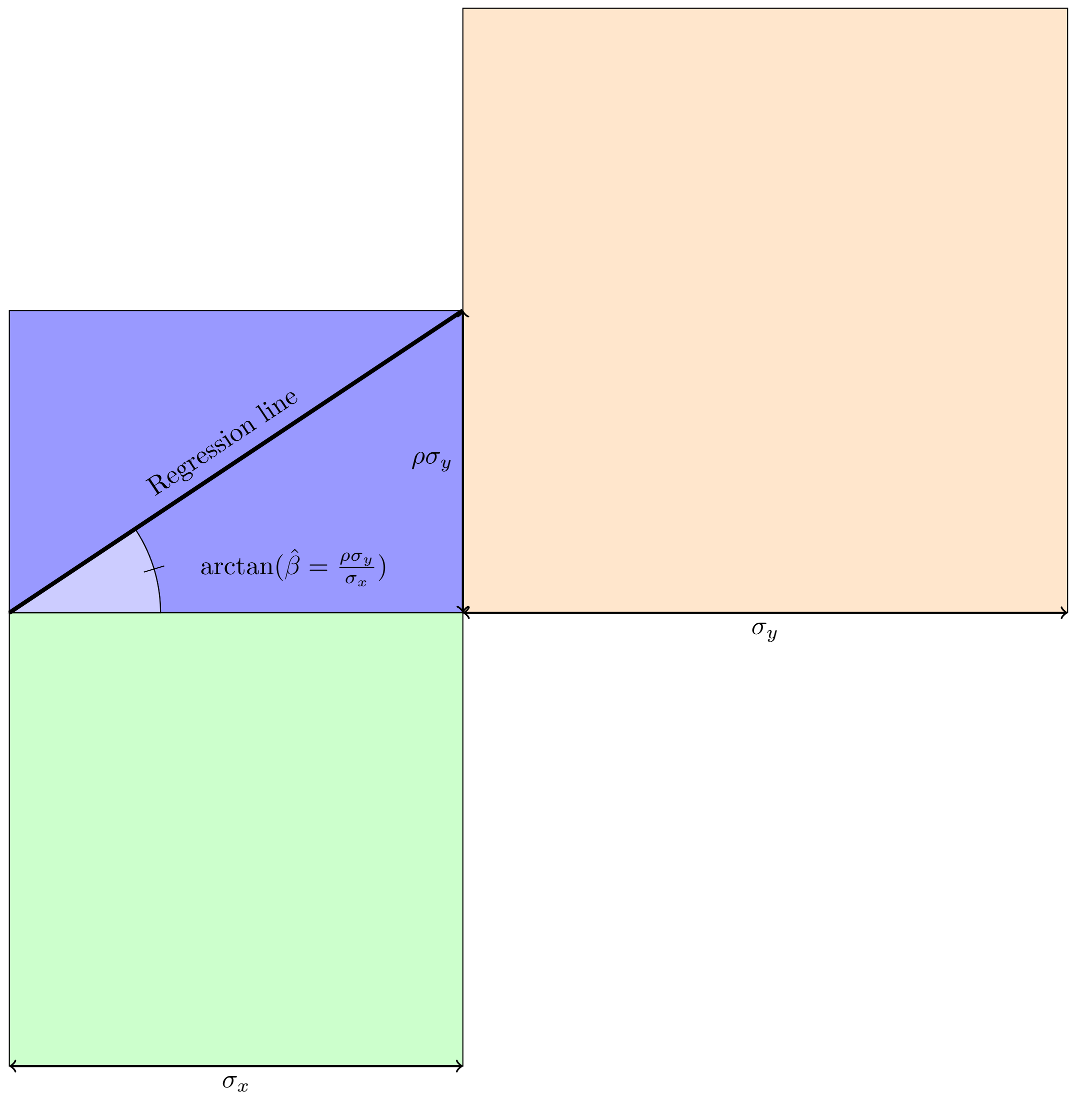
$Var_{max} - Var_{minx} = 2\sigma_X \sigma_Y$

d'où en simplifiant

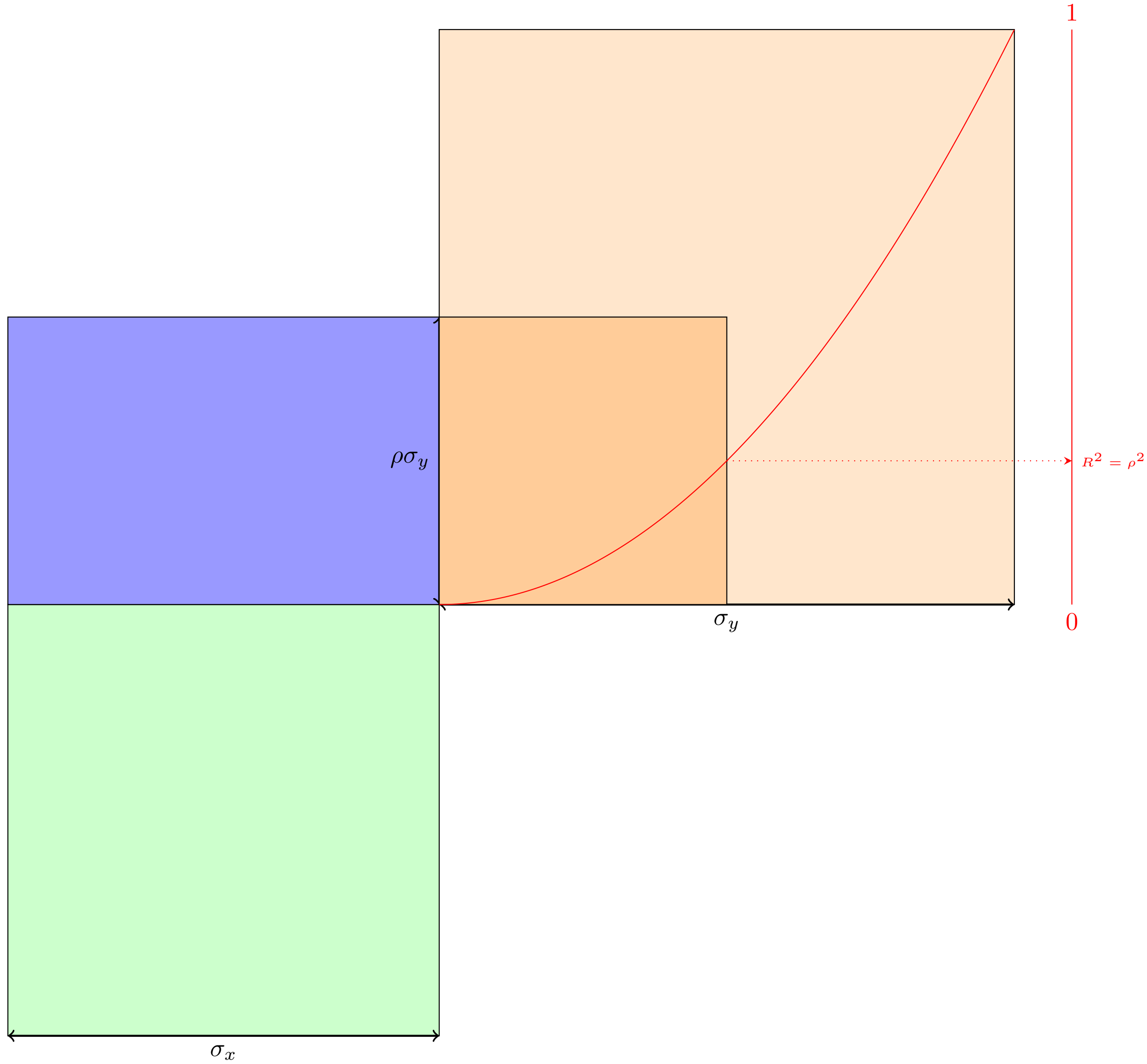
$\rho = \frac{Cov(X,\ Y)}{\sigma_X \sigma_Y}$

1 il faut replacer la définition usuelle du coeff de corrélation et discuter de l'intérêt de notre présentation + intuitive ?
2 donner l'encadrement de var(x+y)en fonction de rho

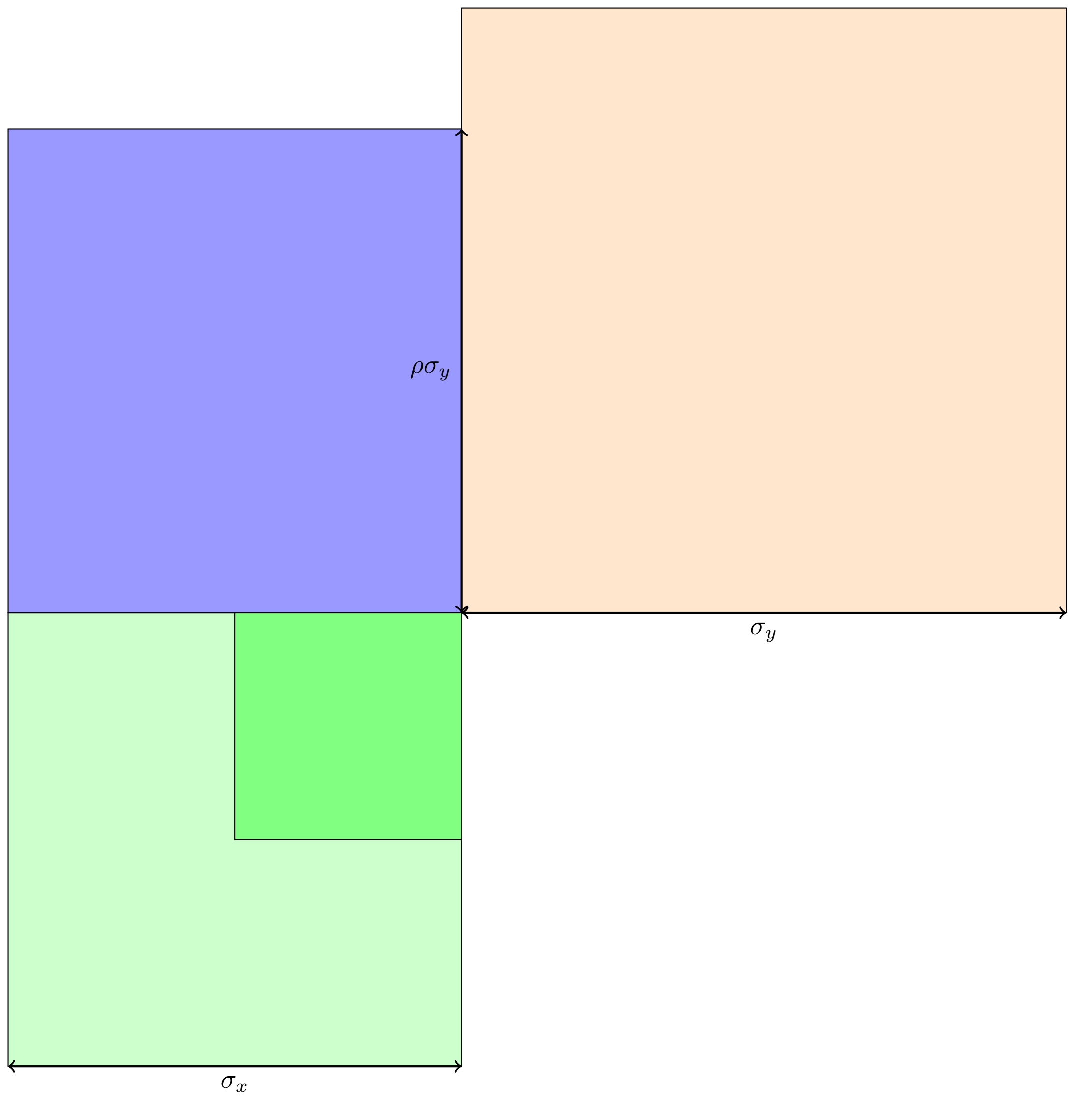
11 LS diagram



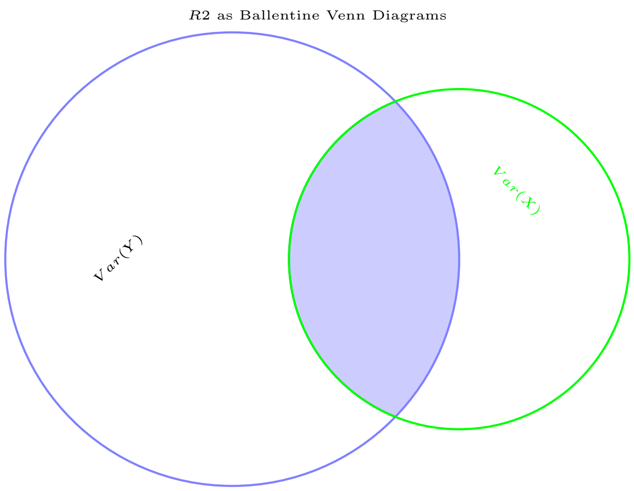
12 LS diagram and R^2



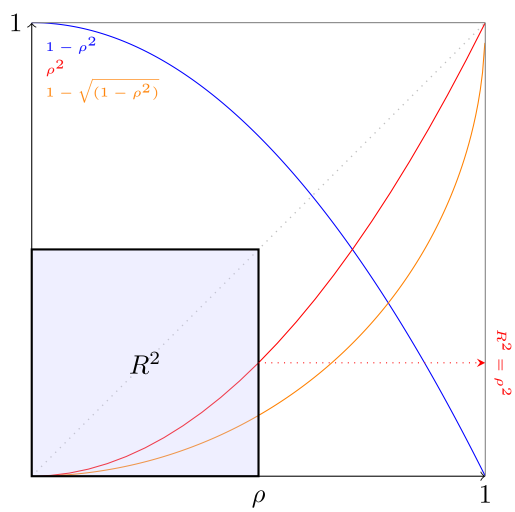
13 Multiple regression partialling out



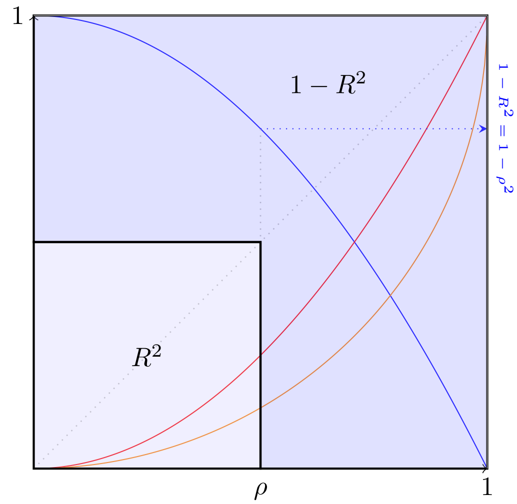
14 Appendix 1: Venn Diagram



15 Appendix 2: explained variance



16 Appendix 3: explained variance and unexplained variance



17 Appendix 4: unexplained variance