



SIG ET DECISION

Rapport d'exercices en statistiques appliquées aux systèmes d'information géographique.

RESUME

Ce projet illustre l'usage des méthodes statistiques et géomatiques pour soutenir la prise de décision. Les analyses sous R et la cartographie des régions marocaines permettent d'évaluer la précision des capteurs, l'impact de campagnes, l'efficacité de traitements et la distribution des indicateurs territoriaux. Un tableau de bord interactif Shiny facilite la visualisation et la priorisation des interventions. Il a été réalisé sous l'encadrement du professeur Rachid EL HALIMI, que nous remercions chaleureusement.

Boris Pengwende SAMNE (Auteur) et Rachid EL HALIMI (Encadrant)

SIG et Prise de décision

Faculté des sciences et techniques de Tanger.
Cycle d'ingénieur Géoinformation 3e Année -
2025/2026 (semestre 5/6)

Table des matières

Exercice 1 – Validation de la précision de deux capteurs GPS	2
Question 1 – Importation et structuration des données	2
Question 2 – Vérification de l'égalité des variances (test de Fisher)	2
Question 3 – Comparaison des précisions moyennes (test de Student)	3
Question 4 – Décision statistique ($\alpha = 4\%$).....	4
Exercice 2 – Évaluation de l'impact d'une campagne de sensibilisation	4
Question 1 – Test de comparaison de proportions ($\alpha = 3\%$).....	4
Question 2 – Intervalle de confiance à 95 % de la proportion après campagne	5
Question 3 – Estimation du taux d'amélioration	6
Exercice 3 – Choix de la dose optimale d'un traitement antiparasitaire	6
Question 1 – Structuration des données.....	7
Question 2 – Vérification des conditions d'application de l'ANOVA.....	7
Question 3 – ANOVA à un facteur	9
Question 4 – Comparaisons multiples (test post-hoc de Tukey).....	9
Exercice 4 : Analyse de l'influence de la filière d'origine sur la réussite	11
Question 1 : Création de la matrice des effectifs	11
Question 2 : Test du Khi-deux d'indépendance	11
Question 3 : Résidus standardisés et identification des cases contributrices.....	12
Question 4 : Décision et interprétation.....	12
Exercice 5 : Cartographie d'indicateurs territoriaux – aide à la décision spatiale	13
Question 1 : Télécharger le shapefile des régions marocaines	13
Question 2 : Importer le fichier avec <code>sf::st_read()</code>	13
Question 3 : Créer un jeu de données attributaire fictif	14
Question 4 : Fusionner les données attributaires avec l'objet spatial	15
Question 5 : Générer une carte choroplèthe avec <code>tmap</code>	16
Question 6 : Décision et interprétation.....	17
Exercice 6 : Tableau de bord interactif pour le suivi d'indicateurs régionaux	18
Question 1 : Reprendre le shapefile et y associer des indicateurs.....	18
Question 2 : Créer l'application Shiny	18
Question 3 : Mise à jour dynamique	20
Question 4 : Décision et interprétation.....	21

Exercice 1 – Validation de la précision de deux capteurs GPS

Mise en contexte

Dans le cadre d'un projet géomatique nécessitant une forte précision de positionnement, un bureau d'études doit choisir entre deux modèles de récepteurs GPS. Des mesures répétées ont été effectuées sur un point fixe afin d'analyser statistiquement les erreurs de positionnement (en mètres). L'objectif est d'utiliser des tests statistiques sous R pour appuyer une décision rationnelle et objectivée.

Question 1 – Importation et structuration des données

Les données d'erreur de positionnement des deux récepteurs ont été importées sous R et organisées dans un `data.frame` comportant deux variables :

- `récepteur` : facteur indiquant le type de récepteur (A ou B),
- `Erreur` : variable numérique représentant l'erreur de positionnement en mètres.

```
> # Données des récepteurs
> A <- c(2.1, 1.8, 2.3, 2.0, 1.9, 2.2, 2.1, 1.7, 2.4, 2.0, 2.2, 1.6,
+       2.3, 1.9, 2.1, 1.8, 2.2, 2.1, 1.7, 2.3, 2.0, 1.8, 2.4, 1.9,
+       2.1, 2.2, 1.7, 2.3, 2.0, 2.1, 1.9)
>
> B <- c(1.5, 1.3, 1.6, 1.4, 1.2, 1.7, 1.5, 1.1, 1.8, 1.4, 1.6, 1.0,
+       1.7, 1.3, 1.5, 1.2, 1.6, 1.4, 1.1, 1.7, 1.3, 1.0, 1.8, 1.4,
+       1.6, 1.5, 1.2, 1.7, 1.3, 1.5, 1.4)
>
> # Création du data.frame
> data_gps <- data.frame(
+   récepteur = factor(rep(c("A", "B"), each = 31)),
+   erreur = c(A, B)
+ )
>
> # Vérification
> str(data_gps)
'data.frame': 62 obs. of 2 variables:
$ récepteur: Factor w/ 2 levels "A","B": 1 1 1 1 1 1 1 1 1 ...
$ erreur   : num 2.1 1.8 2.3 2 1.9 2.2 2.1 1.7 2.4 2 ...
```

Question 2 – Vérification de l'égalité des variances (test de Fisher)

Un test de Fisher (`var.test`) a été réalisé afin de vérifier si les variances des erreurs de positionnement des deux récepteurs peuvent être considérées comme égales.

```

> var.test(erreur ~ recepteur, data = data_gps)

    F test to compare two variances

data: erreur by recepteur
F = 0.95216, num df = 30, denom df = 30, p-value = 0.8941
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
0.4591056 1.9747246
sample estimates:
ratio of variances
0.9521592

```

Interprétation

Le test de Fisher fournit une p-value de **0,8941**, largement supérieure au seuil de risque fixé à $\alpha = 4 \%$. On ne rejette donc pas l'hypothèse nulle d'égalité des variances.

Il est statistiquement justifié de considérer que les deux récepteurs présentent des variances comparables, ce qui autorise l'utilisation du **test de Student avec variances égales** pour la comparaison des moyennes.

Question 3 – Comparaison des précisions moyennes (test de Student)

Un test de Student bilatéral a été appliqué afin de comparer les moyennes des erreurs de positionnement des deux récepteurs, avec un niveau de confiance de **96 %** (correspondant à $\alpha = 4 \%$).

```

> t.test(erreur ~ recepteur,
+        data = data_gps,
+        var.equal = TRUE,
+        conf.level = 0.96)

    Two Sample t-test

data: erreur by recepteur
t = 10.723, df = 60, p-value = 1.425e-15
alternative hypothesis: true difference in means between group A and group B is≠
96 percent confidence interval:
0.4877169 0.7251863
sample estimates:
mean in group A mean in group B
2.035484      1.429032

```

Résultats observés

- Erreur moyenne du récepteur A : **2,035 m**
- Erreur moyenne du récepteur B : **1,429 m**
- p-value : **$1,425 \times 10^{-15}$**
- Intervalle de confiance à 96 % de la différence des moyennes : **[0,488 ; 0,725]**

Question 4 – Décision statistique ($\alpha = 4 \%$)

La p-value obtenue est **nettement inférieure** au seuil de signification de 4 %. L'hypothèse nulle d'égalité des moyennes est rejetée.

Il existe donc une **différence statistiquement significative** entre les erreurs moyennes de positionnement des deux récepteurs.

L'intervalle de confiance à 96 % est strictement positif, ce qui indique que l'erreur moyenne du récepteur A est significativement plus élevée que celle du récepteur B.

Décision finale et recommandation

Dans le cadre d'un projet exigeant une **erreur moyenne inférieure à 2 mètres**, le **récepteur B** est clairement à recommander.

Cette décision est justifiée par :

- une erreur moyenne significativement plus faible (**1,429 m**),
- une différence statistiquement très significative par rapport au récepteur A,
- une validation rigoureuse par tests de Fisher et de Student.

Le récepteur B offre une meilleure précision de positionnement et répond aux exigences du projet.

Exercice 2 – Évaluation de l'impact d'une campagne de sensibilisation

Mise en contexte

Une commune souhaite évaluer l'efficacité d'une campagne de sensibilisation visant à encourager le tri régulier des déchets ménagers. Deux enquêtes ont été réalisées, l'une avant la campagne et l'autre après, afin de comparer les proportions de ménages pratiquant le tri et d'appuyer la décision publique sur une analyse statistique rigoureuse.

Question 1 – Test de comparaison de proportions ($\alpha = 3 \%$)

Un test de comparaison de proportions a été réalisé sous R afin de déterminer si la proportion de ménages triant régulièrement a **augmenté après la campagne**.

Le test est unilatéral et le niveau de confiance est fixé à **97 %**, correspondant à un risque de $\alpha = 3 \%$.

```
> prop.test(x = c(190, 230),
+            n = c(1000, 1000),
+            alternative = "less",
+            conf.level = 0.97)

 2-sample test for equality of proportions with continuity correction

data:  c(190, 230) out of c(1000, 1000)
X-squared = 4.5841, df = 1, p-value = 0.01614
alternative hypothesis: less
97 percent confidence interval:
-1.0000000000 -0.004781959
sample estimates:
prop 1 prop 2
 0.19    0.23
```

Interprétation

La p-value obtenue est **0,01614**, inférieure au seuil de signification de **3 %**. L'hypothèse nulle d'égalité (ou de non-augmentation) des proportions est donc rejetée.
Les résultats montrent que la proportion de ménages triant régulièrement **a significativement augmenté** après la campagne de sensibilisation.

Question 2 – Intervalle de confiance à 95 % de la proportion après campagne

Un intervalle de confiance à **95 %** a été calculé pour estimer la proportion réelle de ménages pratiquant le tri après la campagne.

```
> prop.test(x = 230,
+            n = 1000,
+            conf.level = 0.95)

 1-sample proportions test with continuity correction

data:  230 out of 1000, null probability 0.5
X-squared = 290.52, df = 1, p-value < 2.2e-16
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.2045014 0.2576046
sample estimates:
  p
0.23
```

Interprétation

La proportion estimée après campagne est de **23 %**.
L'intervalle de confiance à 95 % est **[0,2045 ; 0,2576]**, ce qui signifie que la proportion réelle de ménages triant régulièrement se situe très probablement entre **20,5 % et 25,8 %**.

Question 3 – Estimation du taux d'amélioration

Les proportions avant et après campagne ont été comparées afin d'estimer l'amélioration observée.

```
> p_avant <- 0.19
> p_apres <- 230 / 1000
>
> gain <- p_apres - p_avant
>
> p_avant
[1] 0.19
> p_apres
[1] 0.23
> gain
[1] 0.04
```

Résultats

- Proportion avant campagne : **19 %**
- Proportion après campagne : **23 %**
- Gain observé : **0,04**, soit **+4 points de pourcentage**

Décision finale

Au regard des résultats statistiques :

- l'augmentation de la proportion de ménages triant régulièrement est **statistiquement significative**,
- l'amélioration observée est de l'ordre de **4 points de pourcentage**,
- la proportion après campagne est estimée avec précision grâce à l'intervalle de confiance.

La campagne de sensibilisation peut donc être considérée comme efficace.

Elle a eu un impact mesurable et significatif sur le comportement des ménages, justifiant la poursuite ou l'extension de ce type d'actions dans la politique de gestion des déchets.

Exercice 3 – Choix de la dose optimale d'un traitement antiparasitaire

Mise en contexte

Un laboratoire aquacole cherche à déterminer la dose optimale d'un traitement antiparasitaire afin de réduire significativement le nombre de parasites chez le poisson, tout en limitant la quantité de médicament administrée. Plusieurs doses ont été testées expérimentalement et une analyse statistique de type ANOVA est utilisée comme outil d'aide à la décision.

Question 1 – Structuration des données

Les données expérimentales ont été organisées dans un `data.frame` comportant :

- `dose` : facteur indiquant la dose administrée (0, 25, 50, 100, 125 mg),
- `parasites` : variable numérique représentant le nombre de parasites par individu.

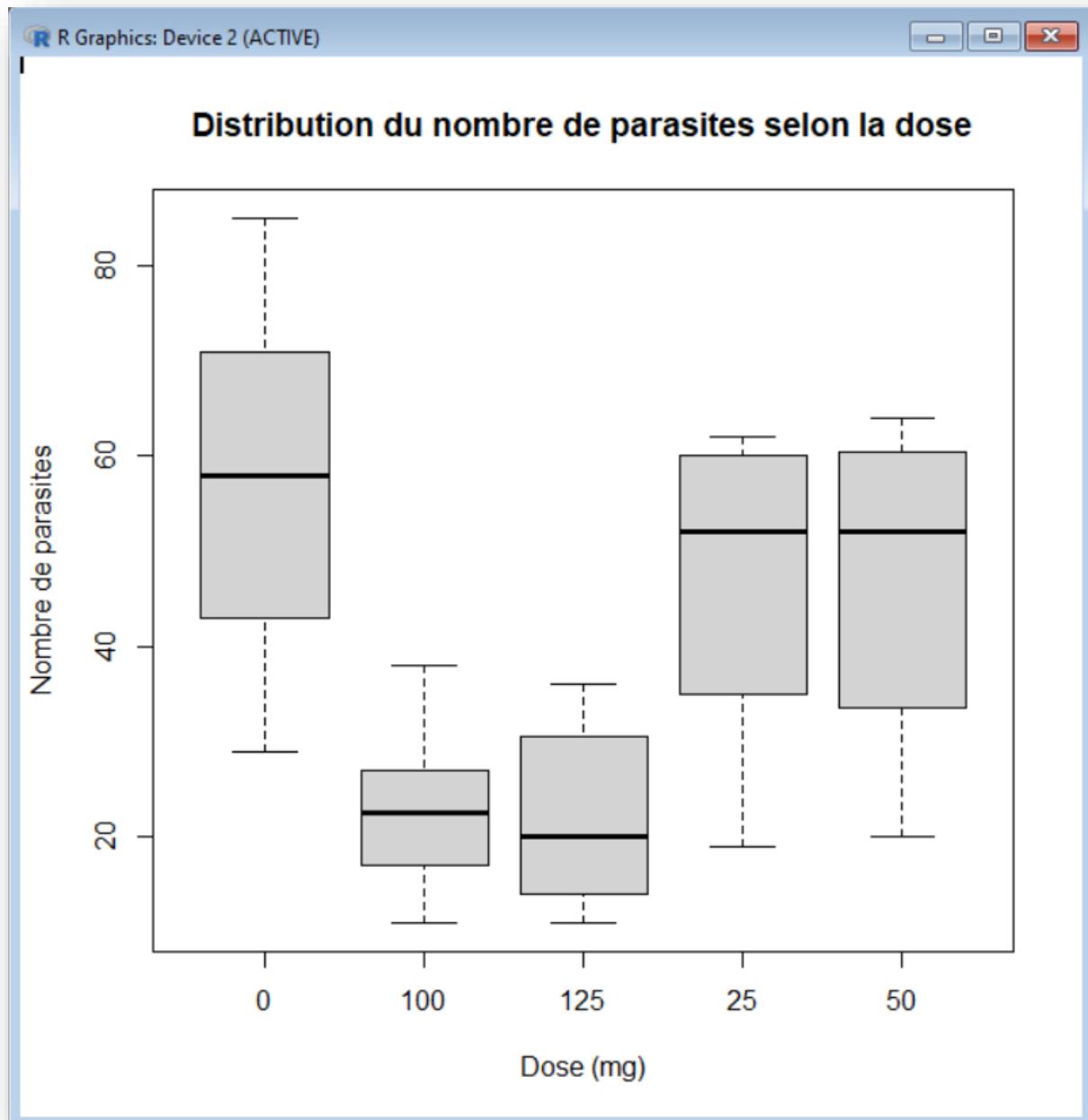
```
> # Création du facteur dose
> dose <- factor(rep(c("0", "25", "50", "100", "125"), each = 12))
>
> # Nombre de parasites
> parasites <- c(
+   50, 65, 72, 46, 38, 29, 70, 85, 72, 40, 57, 59,
+   49, 47, 30, 60, 62, 60, 19, 28, 56, 62, 55, 40,
+   20, 59, 64, 61, 28, 47, 29, 41, 60, 57, 61, 38,
+   20, 23, 38, 31, 27, 16, 27, 18, 22, 12, 24, 11,
+   18, 30, 22, 26, 31, 11, 15, 12, 31, 36, 16, 13
+ )
>
> # Data frame
> data_parasites <- data.frame(dose, parasites)
>
> # Vérification
> str(data_parasites)
'data.frame':   60 obs. of  2 variables:
 $ dose    : Factor w/ 5 levels "0","100","125",...: 1 1 1 1 1 1 1 1 1 ...
 $ parasites: num  50 65 72 46 38 29 70 85 72 40 ...
```

Question 2 – Vérification des conditions d'application de l'ANOVA

2.1 Analyse graphique (boxplot)

Un boxplot a été réalisé afin d'examiner graphiquement la distribution du nombre de parasites en fonction de la dose administrée.

```
> boxplot(parasites ~ dose,
+           data = data_parasites,
+           xlab = "Dose (mg)",
+           ylab = "Nombre de parasites",
+           main = "Distribution du nombre de parasites selon la dose")
```



Interprétation graphique

Le graphique met en évidence une diminution nette du nombre de parasites pour les doses élevées (100 mg et 125 mg) par rapport au témoin. Les doses 25 mg et 50 mg présentent des distributions proches de celle du témoin, suggérant un effet plus limité.

2.2 Normalité des résidus (test de Shapiro-Wilk)

La normalité des résidus du modèle ANOVA a été vérifiée à l'aide du test de Shapiro-Wilk.

```
> modele <- aov(parasites ~ dose, data = data_parasites)
>
> residus <- residuals(modele)
>
> shapiro.test(residus)

  Shapiro-Wilk normality test

data:  residus
W = 0.96363, p-value = 0.07093
```

Interprétation

La p-value obtenue est **0,07093**, supérieure au seuil classique de 5 %.

L'hypothèse de normalité des résidus n'est donc pas rejetée, ce qui valide l'utilisation de l'ANOVA à un facteur.

Question 3 – ANOVA à un facteur

Une analyse de variance à un facteur a été réalisée afin de tester l'effet global de la dose sur le nombre de parasites.

```
> summary(modele)
      Df Sum Sq Mean Sq F value    Pr(>F)
dose      4 12340    3085   17.63 2.3e-09 ***
Residuals 55  9626     175
---
Signif. codes:  0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interprétation

Le test ANOVA donne une statistique **F = 17,63** avec une p-value de **2,3 × 10⁻⁹**, très inférieure au seuil de signification.

On rejette donc l'hypothèse nulle d'égalité des moyennes : **la dose du traitement a un effet statistiquement significatif sur le nombre de parasites.**

Question 4 – Comparaisons multiples (test post-hoc de Tukey)

Un test post-hoc de Tukey a été réalisé afin d'identifier précisément quelles doses diffèrent significativement entre elles.

```
> TukeyHSD(modele)
Tukey multiple comparisons of means
 95% family-wise confidence level

Fit: aov(formula = parasites ~ dose, data = data_parasites)

$dose
    diff      lwr      upr   p adj
100-0 -34.5000000 -49.73204 -19.267963 0.0000004
125-0 -35.1666667 -50.39870 -19.934630 0.0000002
25-0  -9.5833333 -24.81537  5.648704 0.3986122
50-0  -9.8333333 -25.06537  5.398704 0.3723825
125-100 -0.6666667 -15.89870 14.565370 0.9999459
25-100  24.9166667  9.68463 40.148704 0.0002268
50-100  24.6666667  9.43463 39.898704 0.0002656
25-125  25.5833333 10.35130 40.815370 0.0001484
50-125  25.3333333 10.10130 40.565370 0.0001741
50-25   -0.2500000 -15.48204 14.982037 0.9999989
```

Interprétation des comparaisons

- Les doses **100 mg** et **125 mg** montrent une **réduction très significative** du nombre de parasites par rapport au témoin ($p\text{-value} < 0,001$).
- Aucune différence significative n'est observée entre **100 mg** et **125 mg**.
- Les doses **25 mg** et **50 mg** ne diffèrent pas significativement du témoin.
- Les doses élevées (100 mg et 125 mg) sont significativement plus efficaces que 25 mg et 50 mg.

Décision finale

Les résultats montrent que :

- seules les doses **100 mg** et **125 mg** entraînent une réduction significative du nombre de parasites,
- l'efficacité de **100 mg** est statistiquement équivalente à celle de **125 mg**,
- augmenter la dose au-delà de 100 mg n'apporte pas de gain supplémentaire significatif.

La dose de 100 mg est donc recommandée.

Elle permet une réduction significative des parasites tout en **minimisant la quantité de médicament**, ce qui est optimal du point de vue sanitaire, économique et environnemental.

Exercice 4 : Analyse de l'influence de la filière d'origine sur la réussite

Contexte :

L'objectif de cet exercice est d'étudier si la filière d'origine (Économie, Informatique, Statistique) a une influence significative sur la réussite à un examen de géomatique pour 286 candidats. Pour cela, on utilise un **test du Khi-deux d'indépendance** et l'analyse des **résidus standardisés** pour identifier les contributions au rejet de l'indépendance.

Question 1 : Création de la matrice des effectifs

Consigne : Créer une table croisée contenant les effectifs des réussites et des échecs par filière.

Commande R :

```
effectifs <- matrix(c(41, 59, 54, 21, 36, 75), nrow = 2, byrow = TRUE)
rownames(effectifs) <- c("Réussite", "Échec")
colnames(effectifs) <- c("Économie", "Informatique", "Statistique")
effectifs
```

Résultat R :

```
> # Crédation de la matrice
> effectifs <- matrix(c(41, 59, 54, 21, 36, 75), nrow = 2, byrow = TRUE)
> rownames(effectifs) <- c("Réussite", "Échec")
> colnames(effectifs) <- c("Économie", "Informatique", "Statistique")
> effectifs
   Économie Informatique Statistique
Réussite      41          59          54
Échec        21          36          75
```

Question 2 : Test du Khi-deux d'indépendance

Consigne : Vérifier si la réussite dépend de la filière d'origine.

Commande R :

```
test_chi2 <- chisq.test(effectifs)
test_chi2
```

Résultat R :

```
> # Test du Khi-deux
> test_chi2 <- chisq.test(effectifs)
> test_chi2

Pearson's Chi-squared test

data: effectifs
X-squared = 13.828, df = 2, p-value = 0.0009937
```

Question 3 : Résidus standardisés et identification des cases contributrices

Consigne : Calculer les résidus standardisés pour identifier les cellules qui contribuent le plus au rejet de l'indépendance.

Commande R :

```
residus_std <- test_chi2$stdres
residus_std
```

Résultat R :

```
> # Résidus standardisés
> residus_std <- test_chi2$stdres
> residus_std
    Économie Informatique Statistique
Réussite  2.192167   1.975969  -3.685613
Échec     -2.192167  -1.975969   3.685613
```

Question 4 : Décision et interprétation

Consigne : Décider si la filière influence significativement la réussite et identifier les filières favorisées ou défavorisées.

Commande R :

```
test_chi2$p.value
```

Résultat R :

```
> # Décision basée sur la p-value
> test_chi2$p.value
[1] 0.0009936981
```

Interprétation :

- La **p-value = 0.00099** est bien inférieure au seuil de 0,05, ce qui indique que l'on **rejette l'hypothèse d'indépendance**.
- Autrement dit, la réussite à l'examen **dépend de la filière d'origine**.
- L'examen des **résidus standardisés** montre que :
 - La filière **Statistique** a un excès d'échec (résidu standardisé = 3.686) et un déficit de réussite (-3.686), ce qui contribue le plus au rejet de l'indépendance.
 - Les filières **Économie** et **Informatique** présentent des résidus positifs pour la réussite (2.192 et 1.976), ce qui signifie qu'elles sont **relativement favorisées** pour la réussite.

Conclusion : La filière d'origine influence significativement la réussite à l'examen : les étudiants venant des filières Économie et Informatique réussissent davantage que prévu par l'indépendance, tandis que les étudiants de Statistique réussissent moins bien que prévu.

Exercice 5 : Cartographie d'indicateurs territoriaux – aide à la décision spatiale

Contexte :

L'objectif est de produire une **carte thématique des régions du Maroc** illustrant un indicateur de développement, ici la **densité de population**, pour faciliter la prise de décision en aménagement du territoire.

Le shapefile contient 15 régions marocaines avec leurs limites administratives. L'objet `sf` contient 15 entités avec 11 champs, dont `NAME_1` pour le nom des régions.

Question 1 : Télécharger le shapefile des régions marocaines

Consigne : Télécharger le shapefile depuis GADM (niveau 1).

Action : Décompresser le fichier dans un dossier accessible depuis R, ici :

`C:/Users/boris/Downloads/gadm41_MAR_shp`.

Question 2 : Importer le fichier avec `sf::st_read()`

Commande R :

```
library(sf)

# Importer le shapefile
regions_maroc <-
st_read("C:\\\\Users\\\\boris\\\\Downloads\\\\gadm41_MAR_shp\\\\gadm41_MAR_1.shp")
regions_maroc
```

Résultat R :

```

> library(sf)
>
> # Importer le shapefile
> regions_maroc <- st_read("C:\\\\Users\\\\boris\\\\Downloads\\\\gadm41_MAR_shp\\\\gadm41$"
Reading layer 'gadm41_MAR_1' from data source
`C:\\Users\\boris\\Downloads\\gadm41_MAR_shp\\gadm41_MAR_1.shp'
using driver 'ESRI Shapefile'
Simple feature collection with 15 features and 11 fields
Geometry type: MULTIPOLYGON
Dimension:     XY
Bounding box:  xmin: -13.16792 ymin: 27.67007 xmax: -0.9962729 ymax: 35.92264
Geodetic CRS:  WGS 84
> regions_maroc
Simple feature collection with 15 features and 11 fields
Geometry type: MULTIPOLYGON
Dimension:     XY
Bounding box:  xmin: -13.16792 ymin: 27.67007 xmax: -0.9962729 ymax: 35.92264
Geodetic CRS:  WGS 84
First 10 features:
      GID_1 GID_0 COUNTRY          NAME_1 VARNAME_1 NL_NAME_1
1    MAR.1_1   MAR Morocco  Chaouia - Ouardigha     NA     NA
2    MAR.2_1   MAR Morocco  Doukkala - Abda     NA     NA
3    MAR.3_1   MAR Morocco   Fès - Boulemane     NA     NA
4    MAR.4_1   MAR Morocco Gharb - Chrarda - Béni Hssen     NA     NA
5 |  MAR.5_1   MAR Morocco  Grand Casablanca     NA     NA

```

Question 3 : Créer un jeu de données attributaire fictif

Commande R :

```

donnees_regions <- data.frame(
  NAME_1 = regions_maroc$NAME_1,
  population = c(1500000, 1200000, 1000000, 900000, 4000000, 500000,
350000, 2000000, 800000, 1800000, 750000, 600000, 300000, 500000, 250000),
  superficie = c(10000, 9000, 12000, 11000, 800, 70000, 130000, 45000,
15000, 65000, 16000, 14000, 80000, 12000, 50000)
)

# Calcul de la densité
donnees_regions$densite <- donnees_regions$population /
donnees_regions$superficie
donnees_regions

```

```

> donnees_regions <- data.frame(
+   NAME_1 = regions_maroc$NAME_1,
+   population = c(1500000, 1200000, 1000000, 900000, 4000000, 500000, 350000, $,
+   superficie = c(10000, 9000, 12000, 11000, 800, 70000, 130000, 45000, 15000,$
+ )
>
> # Calcul de la densité
> donnees_regions$densite <- donnees_regions$population / donnees_regions$superficie
> donnees_regions
      NAME_1 population superficie   densite
1   Chaoquia - Ouardigha    1500000     10000 150.000000
2   Doukkala - Abda        1200000      9000 133.333333
3   Fès - Boulemane         1000000     12000  83.333333
4 Gharb - Chrarda - Béni Hssen    900000     11000  81.818182
5   Grand Casablanca       4000000      800 5000.000000
6   Guelmim - Es-Semara    500000     70000  7.142857
7 Laâyoune - Boujdour - Sakia El H    350000     130000  2.692308
8 Marrakech - Tensift - Al Haouz     2000000     45000 44.444444
9   Meknès - Tafilalet        800000     15000  53.333333
10   Oriental                1800000     65000 27.692308
11 Rabat - Salé - Zemmour - Zaer     750000     16000 46.875000
12 Souss - Massa - Draâ            600000     14000 42.857143
13   Tadla - Azilal              300000     80000  3.750000
14   Tanger - Tétouan             500000     12000 41.666667

```

Question 4 : Fusionner les données attributaires avec l'objet spatial

Commande R :

```

library(dplyr)

regions_maroc <- regions_maroc %>%
  left_join(donnees_regions, by = "NAME_1")

regions_maroc

```

```

> regions_maroc <- regions_maroc %>%
+   left_join(donnees_regions, by = "NAME_1")
>
> regions_maroc
Simple feature collection with 15 features and 14 fields
Geometry type: MULTIPOLYGON
Dimension:      XY
Bounding box:  xmin: -13.16792 ymin: 27.67007 xmax: -0.9962729 ymax: 35.92264
Geodetic CRS:  WGS 84
First 10 features:
      GID_1 GID_0 COUNTRY                               NAME_1 VARNAME_1 NL_NAME_1
1  MAR.1_1    MAR Morocco           Chaouia - Ouardigha      NA      NA
2  MAR.2_1    MAR Morocco           Doukkala - Abda       NA      NA
3  MAR.3_1    MAR Morocco           Fès - Boulemane       NA      NA
4  MAR.4_1    MAR Morocco          Gharb - Chrarda - Béni Hssen      NA      NA
5  MAR.5_1    MAR Morocco        Grand Casablanca       NA      NA
6  MAR.6_1    MAR Morocco         Guelmim - Es-Semara      NA      NA
7  MAR.7_1    MAR Morocco Laâyoune - Boujdour - Sakia El H      NA      NA
8  MAR.8_1    MAR Morocco Marrakech - Tensift - Al Haouz      NA      NA
9  MAR.9_1    MAR Morocco        Meknès - Tafilalet      NA      NA
10 MAR.10_1   MAR Morocco          Oriental            NA      NA
      TYPE_1 ENGTYPE_1 CC_1 HASC_1 ISO_1 population superficie      densite
1 Region    Region    NA MA.CO    NA     1500000     10000  150.0000000
2 Region    Region    NA MA.DA    NA     1200000      9000  133.3333333

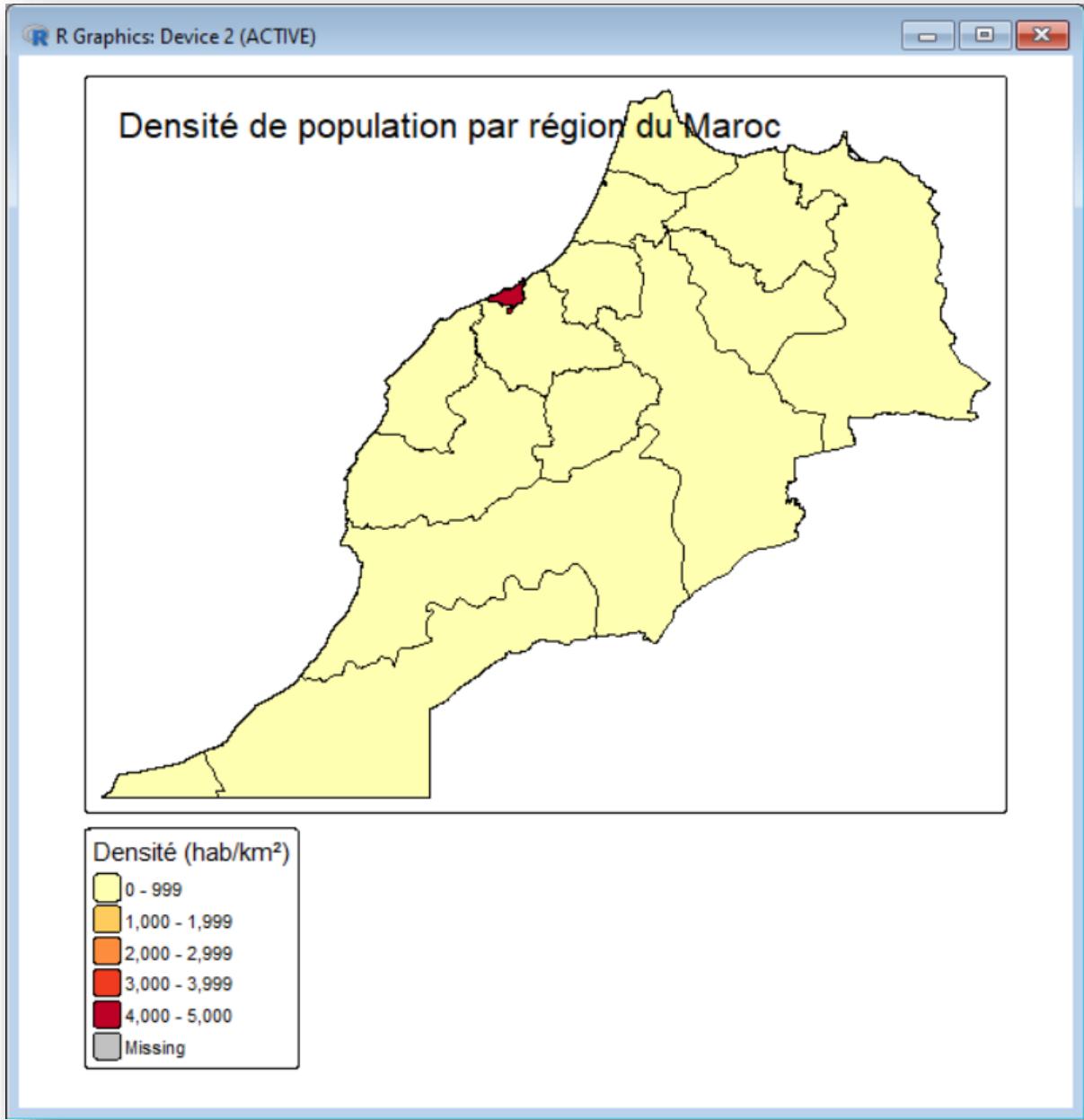
```

Question 5 : Générer une carte choroplète avec tmap

Commande R :

```
library(tmap)

# Carte statique
tm_shape(regions_maroc) +
  tm_polygons("densite", title = "Densité (hab/km2)", palette = "YlOrRd",
  border.col = "black") +
  tm_layout(title = "Densité de population par région du Maroc")
```



Question 6 : Décision et interprétation

Commande R pour consulter les densités :

```
regions_maroc %>% select(NAME_1, densite)
```

Zone pour insérer la capture R ou le tableau de densité

Interprétation :

- Les régions les plus **densément peuplées** (forte densité) sont celles comme **Grand Casablanca**, qui combine une population très élevée sur une petite superficie.

- Les régions à **faible densité** sont généralement les zones désertiques ou montagneuses comme **Laâyoune - Boujdour - Sakia El H** ou **Guelmim - Es-Semara**.
 - **Implications pour l'aménagement du territoire :**
 - Les zones à forte densité nécessitent des infrastructures et services publics adaptés (logement, transports, santé, écoles).
 - Les zones peu peuplées peuvent être ciblées pour le développement rural, touristique ou agricole, pour réduire les déséquilibres territoriaux.
-

Exercice 6 : Tableau de bord interactif pour le suivi d'indicateurs régionaux

Contexte :

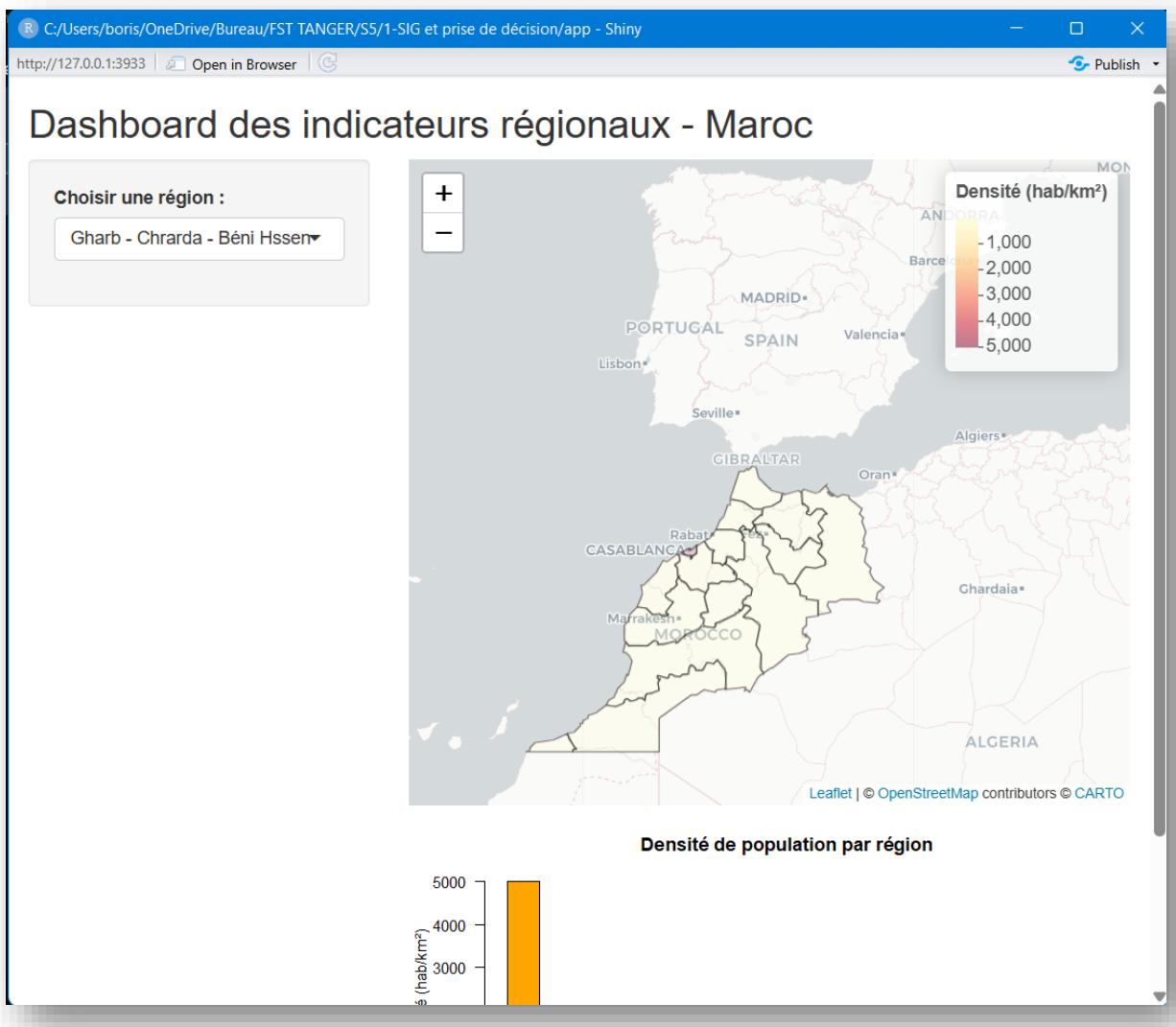
L'objectif est de développer un **dashboard Shiny** permettant à un décideur public de visualiser et filtrer des indicateurs régionaux (population, densité, etc.) pour **prioriser des interventions et planifier l'aménagement du territoire**.

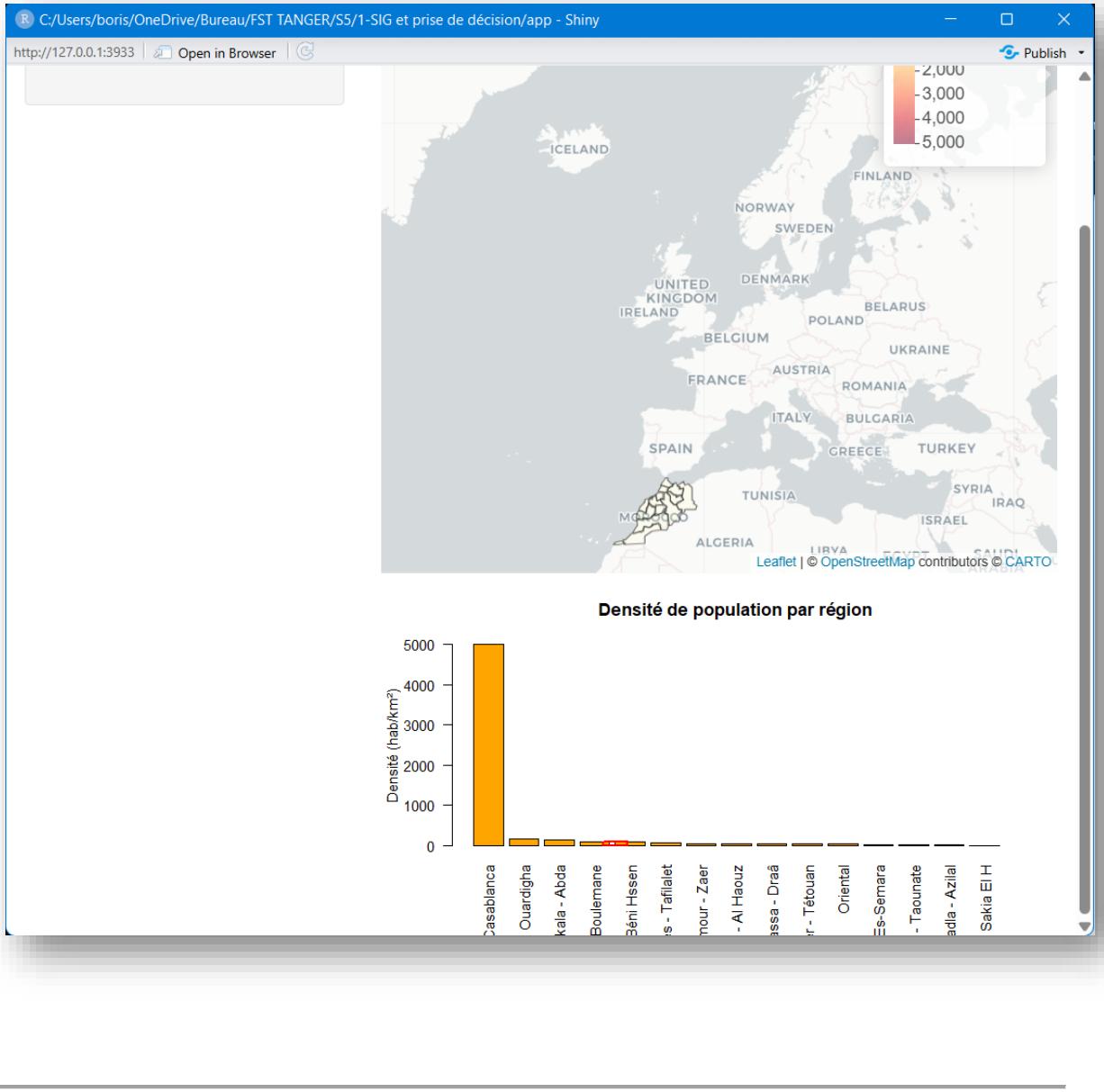
On utilise le shapefile des 15 régions marocaines déjà importé et enrichi avec des indicateurs fictifs ou réels.

Question 1 : Reprendre le shapefile et y associer des indicateurs

Commande R:

Question 2 : Créer l'application Shiny

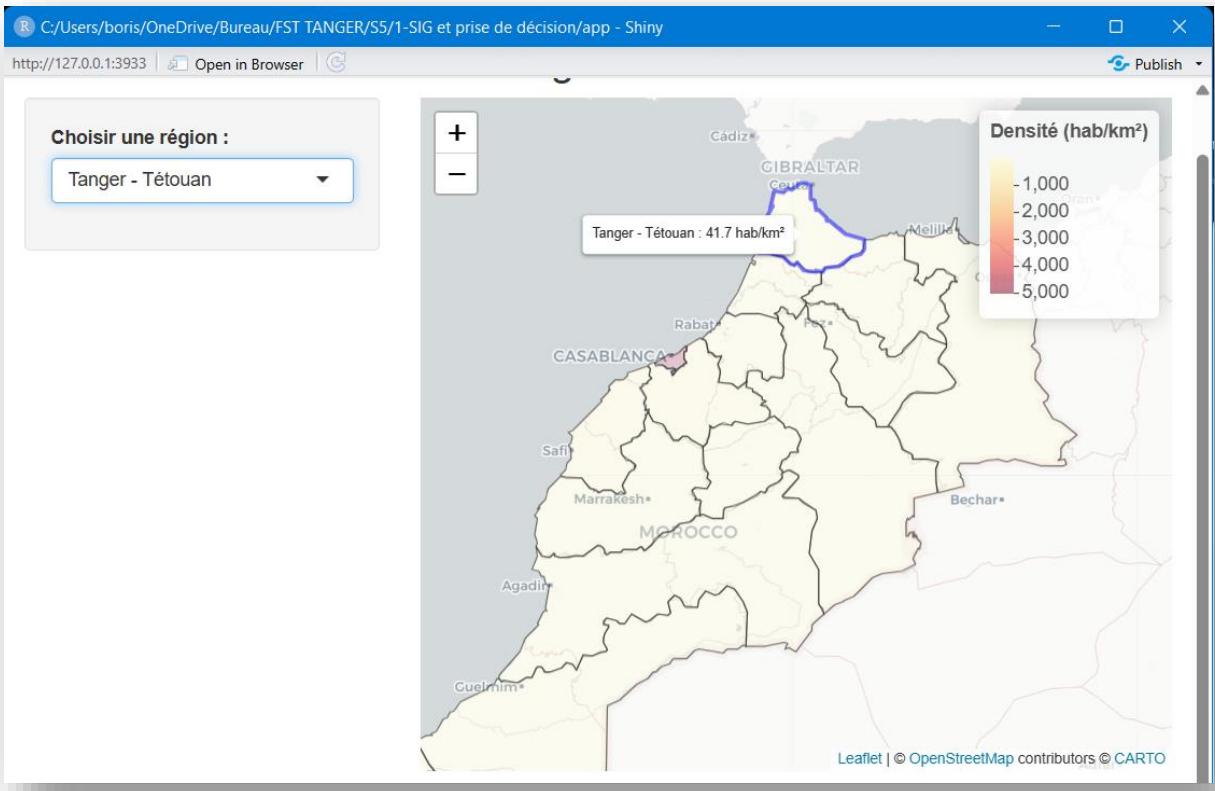




Question 3 : Mise à jour dynamique

Explication :

- La **sélection d'une région** dans le menu déroulant met à jour :
 - La **carte Leaflet** (la région sélectionnée est mise en évidence).
 - Le **diagramme en barres**, avec un contour rouge autour de la région sélectionnée.
- Cela permet une **analyse comparative rapide** de toutes les régions et de leur densité.



Question 4 : Décision et interprétation

Interprétation :

- Le dashboard permet de visualiser rapidement les régions avec forte densité et celles avec faible densité.
- Un responsable de l'aménagement du territoire peut :
 - Prioriser les interventions dans les zones surpeuplées (infrastructures, services publics).
 - Identifier les régions à faible densité pour des projets de développement rural, agricole ou touristique.
 - Comparer facilement les indicateurs entre régions grâce au graphique et à la carte combinés.