**פרויקט קורס ויזואליזציה**

**הרגלי רביצה של פרות**



# מרצה: פיטר בק

רחלי עבו 201431327

אילנה ריכטר 201384724

**תשע"ה סמסטר קיץ**

תוכן עניינים

[1. תקציר 2](#_Toc430165505)

[1.1. תיאור המערכת הנחקרת 2](#_Toc430165506)

[1.2. תיאור המשתמשים 2](#_Toc430165507)

[1.3. מטרות הוויזואליזציה 2](#_Toc430165508)

[1.4. מסקנות 2](#_Toc430165509)

[2. מבוא 2](#_Toc430165510)

[3. תיאור המשתמש 3](#_Toc430165511)

[4. מטרות הוויזואליזציה 4](#_Toc430165512)

[5. תיאור הנתונים 4](#_Toc430165513)

[5.1. כללי 4](#_Toc430165514)

[5.2. איכות נתונים 4](#_Toc430165515)

[5.3. בחירת נתונים 5](#_Toc430165516)

[5.4. רזולוציה 6](#_Toc430165517)

[5.5. טרנספורמציות על הנתונים 7](#_Toc430165518)

[6. דיונים 10](#_Toc430165519)

[6.1. ייצוג קואורדינאטות על ידי מיקום במרחב דו מימדי 10](#_Toc430165520)

[6.2. ייצוג משך השהיה בקואורדינאטה מסוימת 10](#_Toc430165521)

[6.3. יצירת הבחנה בין הפרות השונות 11](#_Toc430165522)

[6.4. יצירת הבחנה בין תת הקבוצות השונות 13](#_Toc430165523)

[6.5. יצירת ממשק אינטראקטיבי 13](#_Toc430165524)

[6.6. כמות המזון שפרה צורכת 14](#_Toc430165525)

[6.7. שימוש ברקע בכדי לייצר הקשר 15](#_Toc430165526)

[6.8. שימוש בפוליגון 15](#_Toc430165527)

[6.9. שימוש במקראיים 15](#_Toc430165528)

[7. מיפוי נתונים למימדים 16](#_Toc430165529)

[8. תוצאות הוויזואליזציה 16](#_Toc430165530)

[8.1. גרפים 16](#_Toc430165531)

[8.2. קוד 17](#_Toc430165532)

[8.3. קבצי עזר 26](#_Toc430165533)

[9. סיכום ומסקנות 26](#_Toc430165534)

[9.1. ערך הוויזואליזציה 26](#_Toc430165535)

[9.2. יתרונות 27](#_Toc430165536)

[9.3. חסרונות 27](#_Toc430165537)

# תקציר

## תיאור המערכת הנחקרת

המערכת הנחקרת היא רפת ניסויית כוללת 32 פרות אשר פעילותן נמדדת בכל עת על ידי חיישנים שונים הממוקמים במרחבי הרפת. עבור כל פרה קיימים המדדים הבאים: משך השהייה בכל קואורדינאטה, כמות המזון שפרה צורכת ומידע רב נוסף, אשר לא נעשה בו שימוש בפרויקט זה.

## תיאור המשתמשים

המשתמש אליו מיועדת הוויזואליזציה הינו חוקר או רפתן השייך למכון אשר מנסה לזהות מגמות שונות בהתנהגות הפרות באמצעות הוויזואליזציה, כמו כן החוקר מנסה לזהות חריגות על ידי התבוננות בוויזואליזציה. וויזואליזציה זו מיועדת למומחה יחיד\מספר מומחים אשר בקיאים בנתונים ובסביבת המחקר.

## מטרות הוויזואליזציה

מטרת העל של הוויזואליזציה הינה לאפשר למתבונן לחקור את התנהגות הפרות ברפת.

1. זיהוי "אזורי רביצה" עבור כל פרה ועבור קבוצת פרות.
2. חלוקה לקבוצות שונות של פרות לפי התנהגות מרחבית ולפי צריכת מזון.
3. זיהוי חריגים- הן בכמות המזון שצורכות הפרות והן בהתנהגותן המרחבית.

## מסקנות

* הפרות מתאפיינות באזורי רביצה דומים יחסית, שעיקרם הוא בקרבת אזור ההאכלה האישי שלהן ובמרכזה של הרפת.
* לא נמצא קשר בין כמות המזון שפרה צורכת לבין ההתנהגות המרחבית שלה.

# מבוא

מינהל המחקר החקלאי הוא המוסד הגדול ביותר בישראל העוסק במחקר חקלאי ונמנה עם הידועים מסוגו בעולם. תפקידיו העיקריים הם לסייע לחקלאי ישראל בפתרון בעיות שונות, לבצע מחקר ופיתוח בנושאים חדשים ומבטיחים בחקלאות ובמדעי המזון ולתכנן, לארגן וליישם מחקר חקלאי בישראל. חוקרי מינהל המחקר החקלאי מלמדים במוסדות להשכלה גבוהה ובקורסים מתקדמים לחקלאות המיועדים למשתתפים מהארץ ומחו"ל ומנחים סטודנטים בעבודותיהם לתארים מתקדמים.

רבים מהחידושים שפותחו במינהל המחקר החקלאי מיושמים כבר ברמה המסחרית בארץ ובעולם. היישום מתבטא במיוחד בתחומים הבאים: גידולים מוגנים (חממות), השקיה, חקלאות באזורים צחיחים, טיפול בתוצרת לאחר הקטיף, הגנת הצומח, מיכון וזנים חדשים של פירות, ירקות וצמחי נוי.

הנהנים הישירים מהישגי מינהל המחקר החקלאי הם החקלאים, התעשיות החקלאיות, ומעל לכל – הצרכנים בישראל, הזוכים למיגון מוצרי חקלאות באיכות טובה.

למינהל רפת מחקרית, אשר ממוקמת במכון וולקני, בצומת בית דגן ליד רחובות. הרפת מרושתת בקולטנים וחיישנים רבים וכך מתאימה לסוגים שונים של מחקרים.

התזה של אילנה ריכטר היא בנושא חקלאות מדייקת, במהלכה תנסה לפתח מודל אשר יחזה כמה אוכל צורכת פרה, וכך ניתן יהיה לחשב בהמשך מדדים רבים כמו רווחיות ויעילות של פרה, בהסתכלות על הפרט ולא על הכלל.

השלב הראשון הוא פיתוח מודל החיזוי ולשם כך יש צורך למצוא כמה שיותר משתנים מסבירים לכמות האוכל אשר צורכת הפרה. לשם כך, החלטנו לייצר וויזואליזציה אשר מאפשרת לחקור את התנהגות הפרות ברפת, בהתחשב בכמות האוכל אשר כל פרה צורכת. ויזואליזציה זו תבוא ראשית לענות על הצורך הנקודתי של המחקר של אילנה- האם קיימת קורלציה בין כמות האוכל שצורכת הפרה ולבין התנהגותה המרחבית, כך שהתנהגות מרחבית יכול לשמש כמשתנה מסביר במודל חיזוי.

בהמשך, בעזרת ויזואליזציה זו, יוכלו חוקרים נוספים וגם העובדים הקבועים ברפת לבחון את ההתנהגות המרחבית של הפרה אל מול כמות המזון שהפרה צורכת, כמו כן הם יוכלו לזהות באמצעות ויזואליזציה זו חריגות, התנהגות של פרה שאינה אופיינית, וכך אולי לזהות מחלות טרם פרצו, פרות "דורשות" ובעיות בשטחי הרביצה של הפרות- בקיץ הפרות נמנעות מלהגיע לאזורים בהם יש יותר מידי שמש, או המאווררים תקולים.

# תיאור המשתמש

* המשתמשים בוויזואליזציה הם חוקרים או רפתנים אשר עובדים ברפת באופן קבוע ומכירים את הנתונים ואת סביבת המחקר בצורה טובה, לאור זאת הסקנו כי אין צורך בהסברים מאוד מפורטים, מלל וכדומה בכדי להסביר את הוויזואליזציה ומספיק מקרא פשוט לצד התרשימים אשר מתאר את מיפוי המשתנים לממדים הוויזואליים.
* למשמש יש סברות שונות בנוגע לנתונים, התפלגותם ותת הקבוצות שהם מכילים, המשתמש מעוניין בחופש פעולה בכדי לחקור תופעות שונות במרחב הנתונים שהוא אף זיהה לפני הסתכלות בוויזואליזציה וכן תוך כדי ההסתכלות בה.
* המשתמש חקר את הנתונים בהיבטים שונים ולכן מכיר וויזואליזציות שונות של הנתונים הללו, כמו כן לאור זאת הוא מסוגל להבין ולפרש וויזואליזציות מורכבות יותר של אותם הנתונים.

# מטרות הוויזואליזציה

המטרה העיקרית שלשמה נוצרה הוויזואליזציה הינה לאפשר לחוקר לזהות מגמות וקבוצות בתוך אוכלוסיית הפרות ולזהות קשר בין ההתנהגות המרחבית של כלל הפרות של קבוצת פרות לבין כמות המזון שהן צורכות.

1. הוויזואליזציה תאפשר זיהוי קל ואינטואיטיבי של "אזורי רביצה" עבור כל פרה בנפרד ועבור קבוצת פרות.
2. הוויזואליזציה תציג חלוקה לקבוצות שונות של פרות לפי התנהגות מרחבית ולפי צריכת מזון.
3. הוויזואליזציה תתמוך בזיהוי קל ומהיר של חריגות, הן עבור פרות ספציפיות והן כמגמה של קבוצת פרות.

# תיאור הנתונים

## כללי

הנתונים אשר נחקרו כוללים מידע לגבי 32 פרות שונות, אשר פעילותן ברפת נמדדת בכל עת על ידי חיישנים שונים:

* **מיקום הפרה, מוצג בקואורדינאטות x, y, z** – מידע זה מתקבל ברזולוציה של סנטימטרים ובמרווחים של אלפיות שנייה.
* **כמות המזון שהפרה צורכת** –לכל פרה עמדת האכלה משלה ורק ממנה היא יכולה לאכול, בכל פעם שכמות המזון בעמדה משתנה (קרי הפרה אוכלת) המערכת מודדת זאת. מידע זה מתקבל ברזולוציה של עשרות גרמים.

עבור כל פרה קיבלנו קובץ אקסל אשר מכיל מידע על משך השהיה של הפרה בכל קואורדינאטה לאורך חודש ימים, כל קובץ כזה הכיל כמה עשרות אלפי רשומות.

כמו כן קיבלנו קובץ אקסל נוסף אשר הכיל מידע לגבי כמות המזון שכל פרה צרכה, גם כן לאורך של חודש ימים, גם קבצים אלו הכילו כמה עשרות אלפי רשומות עבור כל פרה.

## איכות נתונים

* רשומות לא תקינות בקובץ המתאר משך שהיה בכל קואורדינאטה – בבואנו לבחון את הנתונים אשר קיבלנו זיהינו כי ישנם נתונים עבורם משך הזמן שפרה שהתה בקואורדינאטה הינו שלילי, כמו כן זיהינו "קפיצות" במדידת הזמנים עבורם שהתה הפרה בכל קואורדינאטה, החלטנו לנפות את הנתונים הללו מכיוון שכלל לא ניתן לעבוד איתם. ניסינו לחשוב מה ההשלכות של ניפוי זה על הוויזואליזציה. מכיוון שמדובר ברזולוציה מאוד גבוהה אשר מתארת מיקום מדויק מאוד של הפרה, עבור כל קואורדינאטה משך הזמן שהפרה שהתה שם הינו נמוך מאוד (מספר שניות או פחות) ולכן ניפוי זה לא עתיד להטות את הנתונים באופן משמעותי. בפועל, ההשלכה היא שעבור כל פרה נקבל פחות זמן דגימה פר יום, למשל במקום 24 שעות ביממה, נקבל דגימה של 22. חוסר מידע זה לא מטה את הנתונים לכאן או לכאן ולכן אינו מהווה בעיה.
* חוסר במידע – עבור חלק מהפרות היה חסר מידע לגבי כמות המזון שהן צורכות. תחילה חשבנו האם בכל זאת כדאי להשתמש בפרות אלו בוויזואליזציה אך מכיוון שהיה חסר מידע מהותי אשר מהווה בסיס לוויזואליזציה שרצינו לייצר, ומכיוון שלאחר ניפוי הפרות הללו עדיין נשארו לנו 32 פרות מתאימות לייצוג החלטנו להתמקד רק בפרות שעבורן יש לנו מידע מלא.
* רשומות לא תקינות בקובץ המתאר את כמות המזון שהפרה צורכת – בחלק מהנתונים הופיעו "קפיצות" בכמות המזון שנמדדה. מכיוון שמדובר בכמות לא גדולה מידי של רשומות ומכיוון שלא רצינו לוותר על הצגה של פרות נוספות החלטנו לעבור על הנתונים האלו ולתקן את הקפיצות באופן ידני. תיקון זה לא פגע במהימנות הנתונים מכיוון שהרזולוציה בה השתמשנו בסוף הייתה יום שלם והסכימה של כמות המזון שפרה אכלה ביום נשארה כשהייתה.

## בחירת נתונים

* בחינת מסלול ההליכה של הפרה – בתחילת העבודה תהינו האם כדאי להציג את המסלול שהפרות עושות או רק את אזורי הרביצה שלהן. מצד אחד אולי ניתן לזהות דפוס של מסלול של פרה, מצד שני מכיוון שאנו רוצים לזהות אזורי רביצה של הרבה פרות ומתוך כך לפתח מודל לא בטוח שיש טעם בהצגת המסלול. כמו כן הפרות "מטיילות" לאורך היום ברפת והצגת המסלול עלולה להיות מסורבלת ולא ברורה.

בסופו של דבר לאחר בחינת גרפים שונים החלטנו כי לא נציג את המסלול שהפרות עושות אלא רק את האזורים בהן הן שוהות, מתוך כך נוכל לזהות את אזורי הרביצה שלהן. כמו כן החלטנו בשלב זה כי כן נתייחס לזמן ביום אליו הנתונים שייכים- בוקר\ צהריים\ ערב (בהמשך נדון בבחירה זו).

* סינון נתונים – חשבנו כי אולי כדאי לסנן את הנתונים המאוד קצרים בכל קואורדינאטה ולהשאיר רק את אזורי הרביצה של פרות שהינם משמעותיים יותר. מצד אחד זה לא מייצג את כל הנתונים, מצד שני ריבוי הנקודות הקטנות מעמיס על המודל וללא נקודות אלו נוכל לזהות באופן ברור יותר את האזורים שבהן פרות רובצות ולא רק "מטיילות". לאחר שביצענו טרנספורמציות שונות על הנתונים (סכימה לפיx, y - מפורט בהמשך) הצלחנו לייצר תרשים שלעניות דעתנו לא היה עמוס יתר על המידה ולכן החלטנו שלא לסנן את הנתונים הללו.
* מידע לגבי האם הפרה דורשת או לא – תחילה חשבנו לתת התייחסות לנתונים אלו מכיוון שברור לנו שזהו מידע אשר משפיע על התנהגות הפרה ורצינו לבחון כיצד השפעה זו באה לידי ביטוי. כאשר התחלנו לעבוד על הוויזואליזציה ראינו כי יש בידנו מידע רב שחשוב לנו להציג ומכיוון שהשאלה העיקרית הייתה הקשר בין ההתנהגות המרחבית לבין כמות המזון חשבנו כי המידע הזה לא עונה על שאלת הוויזואליזציה ואף מפריע ולכן החלטנו שלא להתייחס אליו.
* קואורדינאטה z – הרפת הינה מישורית, לא קיימים בה הבדלי גובה משמעותיים. לכן החלטנו לוותר על המידע הזה כיוון והוא אינו משתנה ולכן לא משפיע על ההתנהגות המרחבית של הפרה.

## רזולוציה

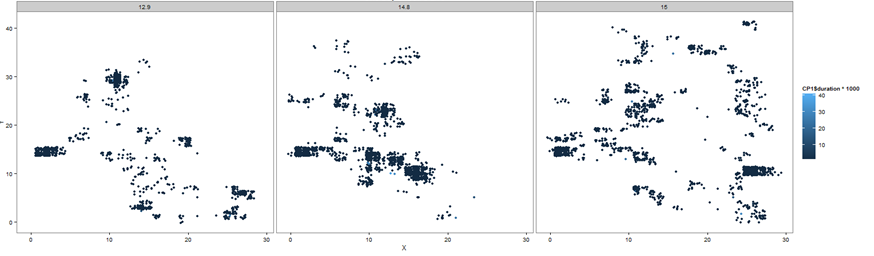
* מיקום במרחב – המידע אשר היה בידנו לגבי מיקום הפרה היה ברזולוציה של סנטימטרים, חשבנו כי אין צורך להשתמש ברזולוציה כזו (גם קשה מאוד לייצוג בהתייחס לגודל הרפת של כמה עשרות מטרים), כמו כן ברזולוציה כזו כמעט ולא תהיה לנו חזרתיות של קואורדינטות. החלטנו כי לעת עתה ננרמל את הנתונים לעשרות סנטימטרים ונראה אם יש לבצע נרמול נוסף בהמשך.

בסיום העבודה על הוויזואליזציה החלטנו כי הנרמול שביצענו אכן מתאים ואין צורך לשנותו.

חלוקת המידע לפי זמן ביום – תחילה חשבנו לייצג את המידע בחלוקה לבוקר, צהריים וערב. זאת מכיוון שידענו כי ברפתות מייחסים חשיבות לזמן ביום ביחס להתנהגות הפרה. לאחר התייעצות עם גורמים ממכון וולקני לגבי החלוקה המתאימה החלטנו לחלק את המידע לפי: בוקר- 00:00-8:00, צהריים- 8:00-16:00 וערב- 16:00-24:00. בחנו את החלוקה הזו עבור פרות שונות ואכן זיהינו הבדלים בהתנהגות הפרות, כצפוי, לפי עומסי החום וזרימת האוויר באזורים השונים (פרה היא חיה האוהבת קור ולכן תחפש תמיד מקומות מוצלים ובעלי יותר אוורור וזרימת אוויר). כך הבנו כי את המידע אשר ניתן להפיק מן החלוקה הזו, כבר יודעים החקלאים והחוקרים, ולכן אין בו צורך.

חלוקה של בוקר\צהריים\ ערב לפי צבע (עבור פרה אחת):

****

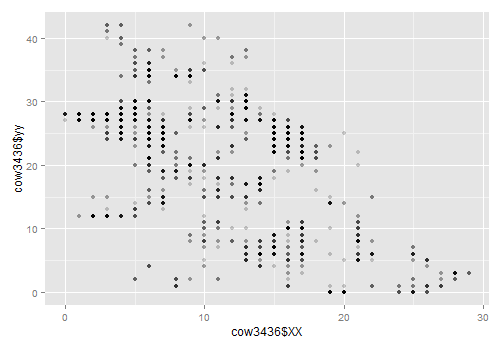
****חלוקה של בוקר\צהריים\ ערב ל-3 גרפים שונים (עבור פרה אחת):

## טרנספורמציות על הנתונים

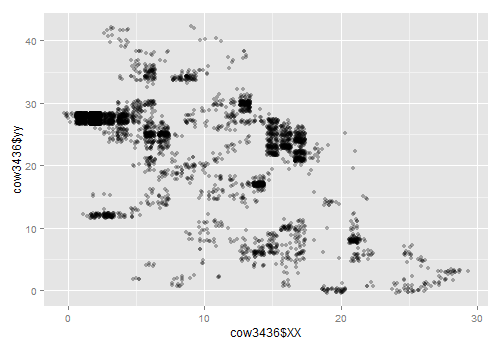
לאחר שבחנו את הנותנים עלו מספר שאלות לגבי אופן ייצוגם:

* חישוב משך שהייה כולל בנקודה במרחב – מכיוון שהחלטנו שלא לייצג את מסלול ההליכה של הפרה במרחב אלא רק את הנקודות בהן היא נמצאת נוצר מצב שבו יש נקודות רבות בכל קואורדינאטה אשר מסתירות אחת את השנייה. אפילו כאשר השתמשנו בשקיפות מידע רב הלך לאיבוד.

גרף המתאר אזורי שהייה של פרה אחת:



על מנת שלא לאבד מידע ניסינו להשתמש ב"Jitter":

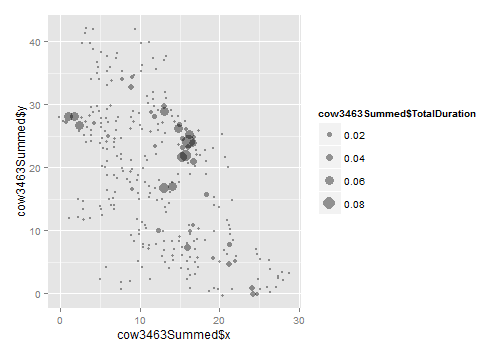


אכן הצלחנו לקבל תוצאות טובות יותר, אבל הגרף היה נראה עמוס עוד לפני שהוספנו את שאר המידע אשר רצינו לייצג.

החלטנו כי מכיוון שוויתרנו על הרזולוציה של מסלול ועל החלוקה לבוקר, צהריים, ערב, ומכיוון שכרגע קיימות רשומות רבות שמייצגות תזוזה בסנטימטרים בעוד אנו החלטנו לעבור לרזולוציה של עשרות סנטימטרים, נבצע סכימה של משך השהייה הכולל בכל קואורדינאטה ליום, כלומר עבור כל קואורדינאטה x, y סכמנו את משך הזמן הכולל בה הפרה שהתה בה במשך היום.

מצורף מטה גרף המתאר עבור כל קואורדינאטה את משך השהייה הכולל בה של פרה אחת בלבד.

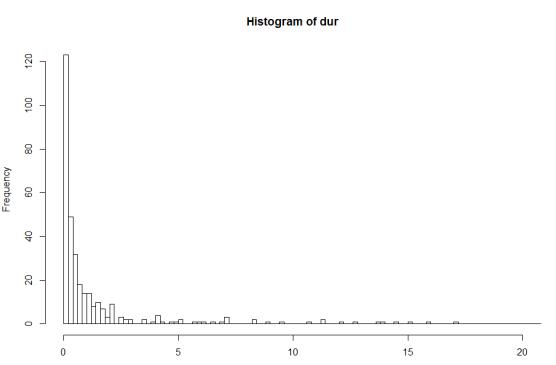
משך ממופה לגודל:



לאחר בחינת הגרפים השונים החלטנו כי זו הדרך בה נציג את המידע שבחנו עד כה.

* התפלגות של משך שהייה בקואורדינאטה מסוימת – לאחר שהחלטנו לאחד את משך הזמן של כל פרה בכל קואורדינאטה החלטנו לבחון כיצד מידע זה מתפלג במטרה להחליט כיצד לייצגו.

היסטוגרמה של משך שהיה בקואורדינאטה:



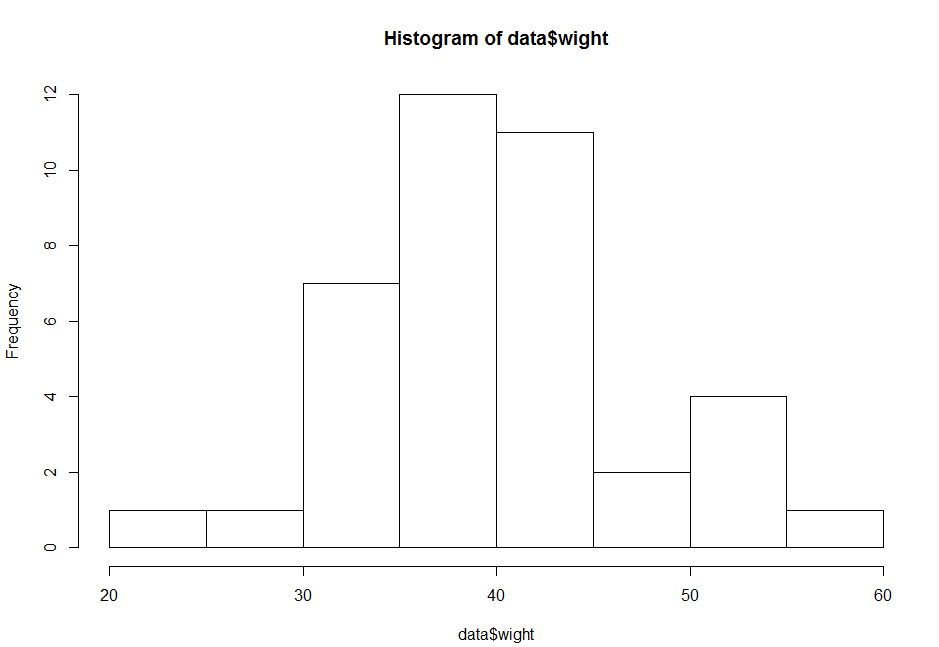
ניתן לראות כי משכי השהייה בכל קואורדינאטה מתפלגים (בקירוב) באופן לוגריתמי, אי לכך החלטנו לייצג את משכי השהייה על ידי שימוש בלוגריתם של הזמן.

הסבר על אופן החישוב:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Value** | **Value in hours** | **log(x)** | **R(x)** |
| minVal | 1.30E-05 | 00:00:01 | -11.2505612 | 0.000013 |
| maxVal | 0.160066 | 03:50:30 | -1.83217155 | 1.000013 |
| other value | 0.000118 | 00:00:10 | -9.04550413 | 0.394201 |

* התפלגות כמות המזון שהפרה צורכת – הרעיון הראשוני היה לייצג את כמות המזון שפרה צורכת על ידי צבע או פוליגון, לשם כך היינו צריכות לייצר משתנה קטגוריאלי שמצביע על השתייכות לקבוצה מסוימת.

על מנת לבחון כיצד מתפלגת כמות המזון שפרה צורכת ביצענו היסטוגרמה:



ניתן לראות כי הנתונים מתפלגים בקירוב באופן נורמאלי לכן אין צורך לבצע טרנספורמציה כלשהי עליהם. עם זאת החלטנו כי נחלק את הנתונים לחמש קבוצות שוות גודל וזאת על מנת לייצג את ההשתייכות לתת קבוצה כלשהי, אשר בהמשך תמופה למימד בוויזואליזציה.

# דיונים

במהלך עבודתנו, תיעדנו את הדיונים השונים שהיו לנו על האופן המתאים ביותר לממש את הוויזואליזציה, כך שתענה על כל המטרות שהצבנו בצורה האופטימאלית והנוחה ביותר.

## ייצוג קואורדינאטות על ידי מיקום במרחב דו מימדי

המהות של נתוני הקואורדינאטות הינו מיקום במרחב ולכן חשבנו שיהיה אינטואיטיבי למפות זאת למיקום במרחב הוויזואליזציה. ניסינו לחשוב האם זוהי המחשבה האינטואיטיבית ביותר או אם יש מימד מתאים יותר לייצוג זה, אך על בסיס הדיונים שערכנו בהקשר זה בכיתה, ועל בסיס העובדה שאם לנו זה היה הרעיון המידי, כך גם לצופה מיפוי זה יהיה אינטואיטיבי ופשוט להבנה, בחרנו באופציה זו.

## ייצוג משך השהיה בקואורדינאטה מסוימת

מכיוון שהחלטנו כי נציג את משך השהייה של הפרות בכל קואורדינאטה ולא את מסלולי ההליכה של הפרות נוצרה דילמה כיצד להציג נתונים אלו באופן שיהיו ברורים ויאפשרו זיהוי מגמות ואזורי רביצה ומצד שני לא יעמיסו יתר על המידה, בכל זאת מדובר בהרבה מידע עבור כמות גדולה של פרות.

חשבנו על שתי אפשרויות:

* הראשונה היא הצגת המידע על ידי "פוליגון", כאשר גודל הפוליגון ייצג את משך השהייה של פרה בנקודה (או לחלופין את המשך השהייה של כל הפרות בקואורדינאטה, אך את זה חשבנו שכדאי שנחליט בשלב מתקדם יותר, לאחר שנבחן תרשימים שונים).
* האופציה השנייה היא שימוש ב"מפת חום" שבעזרתה נציג את אזורי הרביצה של הפרות, מצד שני לא היינו בטוחות כיצד ניתן לייצג באופן זה גם את מספר הפרות וגם את משך השהייה שלהן בכל אזור בדרך שתהיה ברורה.

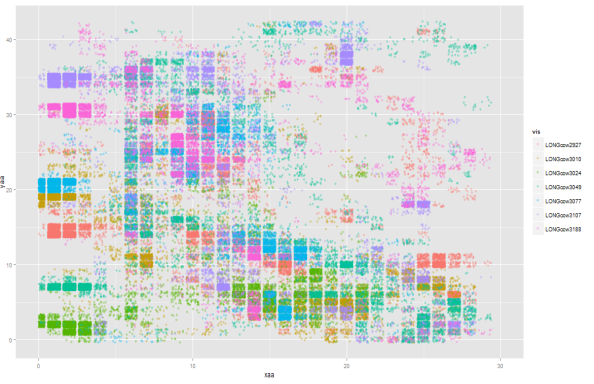
החלטנו כי ייצוג משך השהייה בעזרת פוליגון יהיה מתאים יותר לוויזואליזציה אשר ברצוננו לייצר מכיוון שמפת חום הייתה מאלצת אותנו לוותר על מידע נוסף אשר ברצוננו לייצג.

## יצירת הבחנה בין הפרות השונות

התכנון הראשוני היה להציג על תרשים אחד את כלל הפרות שחקרנו, ולייצר הבחנה בניהן על ידי פוליגון או על ידי צבע.

ניסינו זאת תחילה על מספר מצומם של פרות, בשביל להבין כיצד זה יראה.

תרשים אשר מציג שבע פרות יחד:



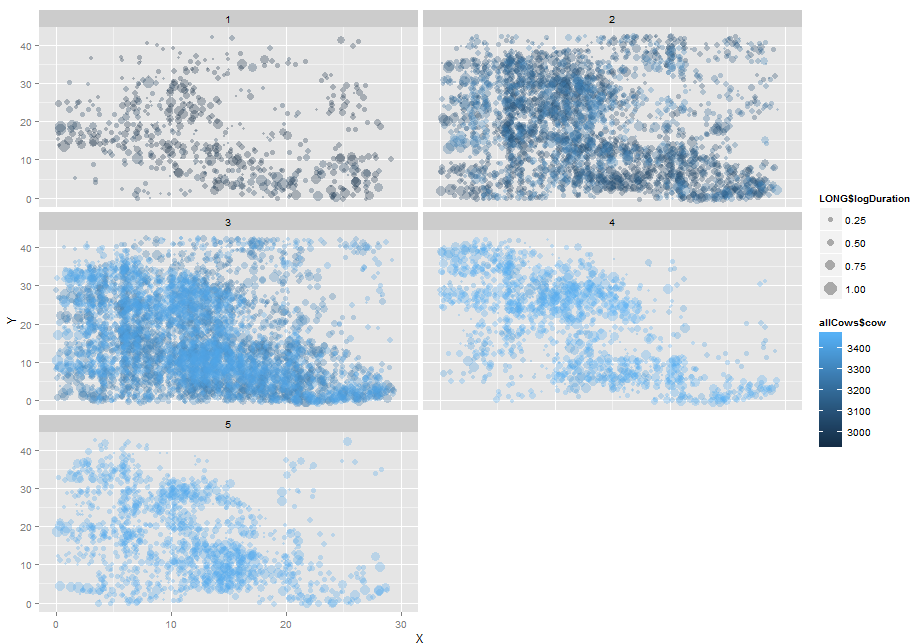
ניתן לראות כי כבר בהוספת הצבע נוצר עומס ויזואלי גבוהה מאוד.

בעיה נוספת בה נתקלנו היא שצבע או סוג פוליגון מתאימים לייצוג של סדר גודל של 6 אובייקטים נפרדים (מקסימום 12) וכי לנו יש 32 פרות שעלינו לייצג.

לאחר בחינת המצב הבנו כי נוכל לייצג את המידע שאנו רוצות להציג בשתי דרכים בלבד:

* סכמת הזמן הכולל בקואורדינאטה עבור כל הפרות יחד – זהו רעיון שעלה כבר בשלב מוקדם יותר של הפרויקט אך משום שמטרת הוויזואליזציה הייתה הצגת ההבדלים בין כל פרה ופרה הבנו כי הוא אינו עונה על שאלת הוויזואליזציה כלל ואינו מתאים לצרכינו.
* פיצול הוויזואליזציה למספר תרשימים אשר כל תרשים מציג פרה או קבוצת פרות נפרדת:
  + קבוצת פרות – חשבנו תחילה להציג תרשים עבור כל קבוצה בחלוקה לפי כמויות המזון שהפרות אוכלות.

תרשים ראשוני אשר מציג חלוקה כזאת:

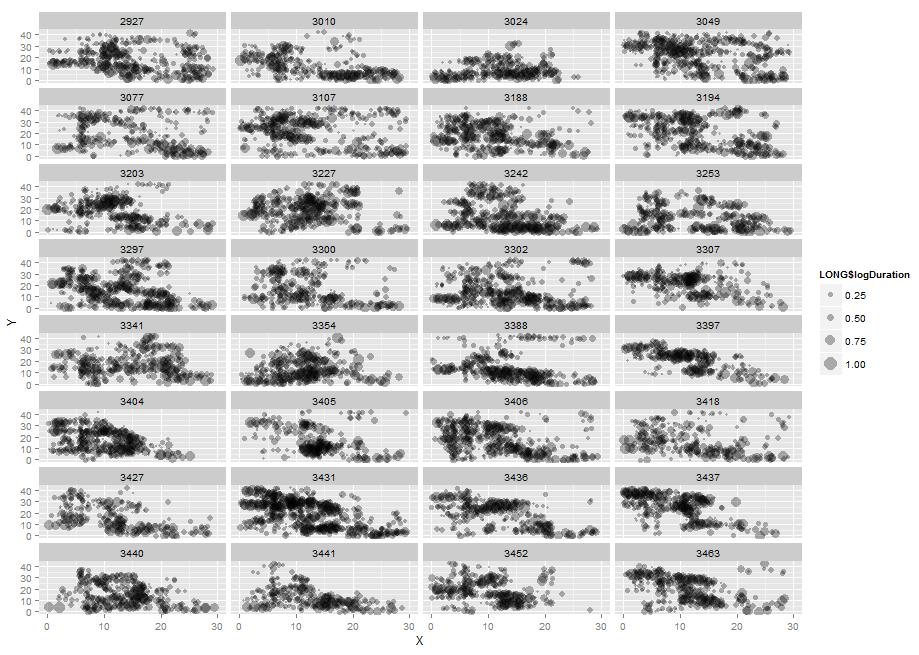


ניתן לראות כי כבר בחלוקה כזאת יש עומס רב מידי בכל תת תרשים וכי לא ניתן להפיק ממנו את המידע הדרוש עבור כל פרה.

חשבנו לסכום עבור כל קואורדינאטה X, Y את המשך הכולל של הפרות באותה תת קבוצה אך רעיון זה נפסל מהסיבות שצוינו לעיל (ברצוננו לייצג את ההתנהגות של כל פרה לכשעצמה). סיבה נוספת שבגינה פסלנו את הרעיון היא שמכיוון שההתפלגות צריכת המזון הינה נורמאלית מספר הפרות בכלל תת קבוצה של אינו שווה (ניתן לזהות זאת גם בגרף הנוכחי) ולכן נוצרת הטיה בתצוגה כזאת, חלק מהתרשימים מכילים כמות גדולה יותר של פרות מאחרים.

* + חלוקה לתרשים נפרד עבור כל פרה – מכיוון שהיה חשוב לנו שהצופה יזהה את השוני בין הפרות בצורה טובה ויוכל להבחין באזורי הרביצה עבור כל פרה החלטנו שחלוקה כזו תהיה טובה ותתאים לצורכי הוויזואליזציה.

תרשים ראשוני של חלוקה לפי פרה:



החלטנו כי אכן הצגת תרשימים שונים בחלוקה לפי פרה תתאים לצורכי הוויזואליזציה של זיהוי אזורי רביצה עבור כל פרה.

## יצירת הבחנה בין תת הקבוצות השונות

משהחלטנו להציג כל פרה בתרשים נפרד היה צורך להחליט כיצד נייצג את תת הקבוצות לפי צריכת מזון. חשבנו על שתי אפשרויות עיקריות:

* פוליגון – הבעיה העיקרית עם פוליגון הוא שקשה לייצר ממנו סקאלה כלשהי שנותנת מידע מעבר לחלוקה לקבוצות (לא ניתן להבין האם עיגול הוא "יותר" ממשולש) ולכן החלטנו לוותר כל אפשרות זו.
* צבע – בעזרת צבע ניתן יהיה גם לייצג הפרדה ברורה בין תת הקבוצות וגם "לסמן" לצופה תרשימים חריגים שאליהם עליו להתייחס. (בחירת פלטת הצבעים מפורטת בהמשך).

## יצירת ממשק אינטראקטיבי

אמנם חלוקת הוויזואליזציה לגרפים שונים עבור כל פרה מאפשרת הסתכלות כללית טובה, אך בהצגה זו אנו מאבדות את האפשרות לתחקר כל פרה\קבוצת פרות מכיוון שכל גרף הוא קטן למדי. על מנת לתת מענה לבעיה זו עלה הרעיון של להשתמש ב "Shiny" , שימוש זה יאפשר לנו תצוגה אינטראקטיבית אשר מצד אחד מציגה הסתכלות כללית על כל הפרות ומצד שני מאפשרת תחקור כל קבוצה ואפילו כל פרה בנפרד בתצוגה גדולה וברורה יותר.

מכיוון שההנחה ההתחלתית שלנו היא שהוויזואליזציה מיועדת לחוקר\רפתן\קבוצת חוקרים מצומצמת החלטנו כי נכון יהיה לאפשר לאלה חופש פעולה בהסתכלות על הוויזואליזציה ולא להציג תמונה אחת קבועה ללא דרגות חופש.

הרעיון הראשוני היה להציג גרף כללי של כל הפרות כפי שפורט לעיל ובנוסף לאפשר למשתמש לסנן קבוצות (לפי כמות אכילה). הוספנו לממשק את האפשרות להוסיף קווים אשר מאפשרים זיהוי מהיר של אזורי הרביצה העיקריים.

כמו כן הוספנו אפשרות לתחקר גרף של פרה אחת בלבד זאת ע"י הצגת הגרף עבור פרה אחת ותצוגה נוספת אשר מאפשרת תחקור תת אזורים.

## כמות המזון שפרה צורכת

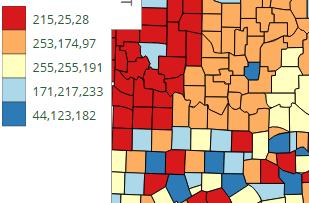
חשבנו כי הדרך מתאימה לייצג את כמות המזון שהפרה אוכלת בכדי שבאמת תיתן מידע למשתמש תהיה בעזרת "סקאלה" בין שני צבעים כאשר כל קיצון מייצג האם הפרה אוכלת "מעט" או "הרבה".

על ידי ביצוע היסטוגרמה לנתונים זיהינו כי נתון זה מתפלג נורמאלית (בקירוב) עם ערך ממוצע של 40 ק"ג ליום.

לאחר שהחלטנו על שימוש אפשרי בסקאלת שני צבעים ,(divergent) תהינו אילו שני צבעים מתאימים לייצוג. מצד אחד למדנו בכיתה כי ישנם צבעים שהקונטרסט ביניהם הינו קל יותר לזיהוי וגם ניתן בעזרתו לתת מענה לעיוורון צבעים וגם בכלל לאפשר הבחנה טובה יותר בין הקבוצות, מצד שני חשבנו שאת היחס של ה"מעט-הרבה" אולי כדאי לייצג בצבעים שמתאימים להקשר. הרי עבור החוקר/הרפתן מידע על פרות שאוכלות הרבה או מעט, הוא משמעותי וראוי לייצגו באופן בולט.

השתמשנו באתר ColorBrewer על מנת לבחון פלטות צבעים שונות. לבסוף החלטנו להשתמש בפלטה של 5 צבעים (כמספר הקבוצות) אשר נעה בין אדום (=הפרה אוכלת מעט) לכחול (=הפרה אוכלת הרבה).

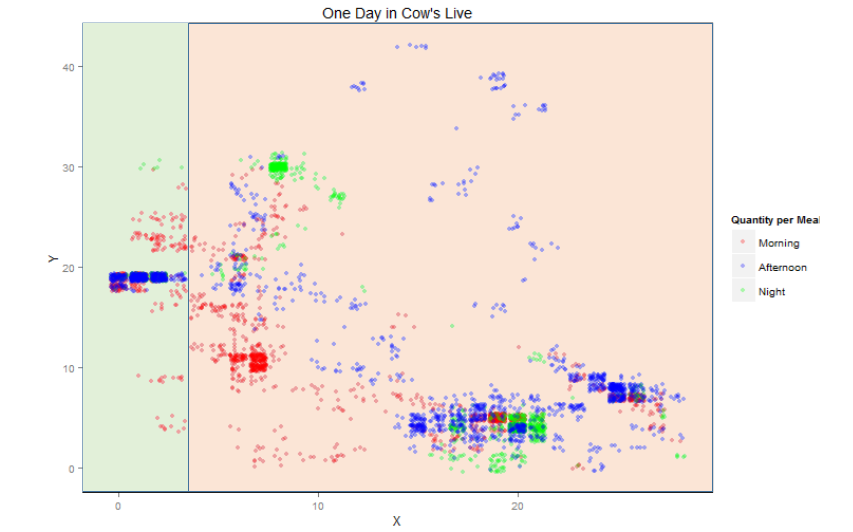
הבחירה להשתמש בפלטה של חמישה צבעים היא על מנת שכלל הפרות שאינן "חורגות" מהערכים הרצויים ייוצגו באמצעות צבע אמצעי, שאינו נוטה לאף כיוון של הסקלה.



## שימוש ברקע בכדי לייצר הקשר

חשבנו שאולי בכדי לתת לחקלאי הקשר של אזורי הרביצה של הפרות כדאי להשתמש ברקע שמייצג מה למעשה נמצא באזורים הללו. (למשל היכן נמצאת עמדת ההאכלה). חשבנו שאולי כדאי לייצג זאת בצורה כללית ביותר על ידי רקע בהיר ואף לתת למשתמש בחירה האם להשתמש במרקע או לא.

לאחר מספר ניסיונות הבנו כי הוספת רקע אמנם מוסיף מידע לצופה לגבי מהו אזור ההאכלה של כל פרה, אך יותר מכך שמתווסף מידע מתווסף גם עומס לוויזואליזציה (עמדת ההאכלה נמצאת בצד השמאלי של התרשים, היכן שגם מופיעים פוליגונים רבים). אי לכך החלטנו שנוותר על השימוש ברקע.

דוגמא עם רקע, בבחינת האפשרות:

## שימוש בפוליגון

ניסינו פוליגונים שונים על מנת לבדוק איזה מהם המתאים ביותר לוויזואליזציה, מכיוון שהחלטנו לא להשתמש בצורת הפוליגון לייצוג מידע כלשהו היה מדובר רק בבחירת הפוליגון שיתאים ביותר מבחינה ויזואלית.

לבסוף החלטנו שעיגול יהיה הפוליגון הפשוט והאינטואיטיבי ביותר ולכן השתמשנו בו.

## שימוש במקראיים

ניהלנו דיון לגבי השימוש במקראיים השונים בתצוגות שלנו.

בגרף הראשי, החלטנו להשאיר מקרא אחד בלבד, אשר מסביר מהם הצבעים הנראים. הסיבה שבגללה לא השארנו גם מקרא לגודל העיגולים הוא כיוון ובתצוגה שכזאת לא ניתן כלל לראות את ההבדלים בגדלי העיגולים השונים.

בגרפים האחרים כן ניתן לראות את הבדלי הגדלים ולכן השארנו את המקרא הזה.

כפי שמפורט במקומות המתאימים בדו"ח זה, גדלי העיגולים נקבעים לפי משך השהייה של הפרה בקואורדינטה הספציפית, והם עוברים מניפולציה לוגריתמית כפי שנלמד בקורס. אמנם לעין שלנו קשה להשוות את גדלי העיגולים למקרא, אך אין המטרה בזיהוי המשך המדויק אלא בהבנת סדר הגודל, האם העיגול גדול או קטן באופן יחסי.

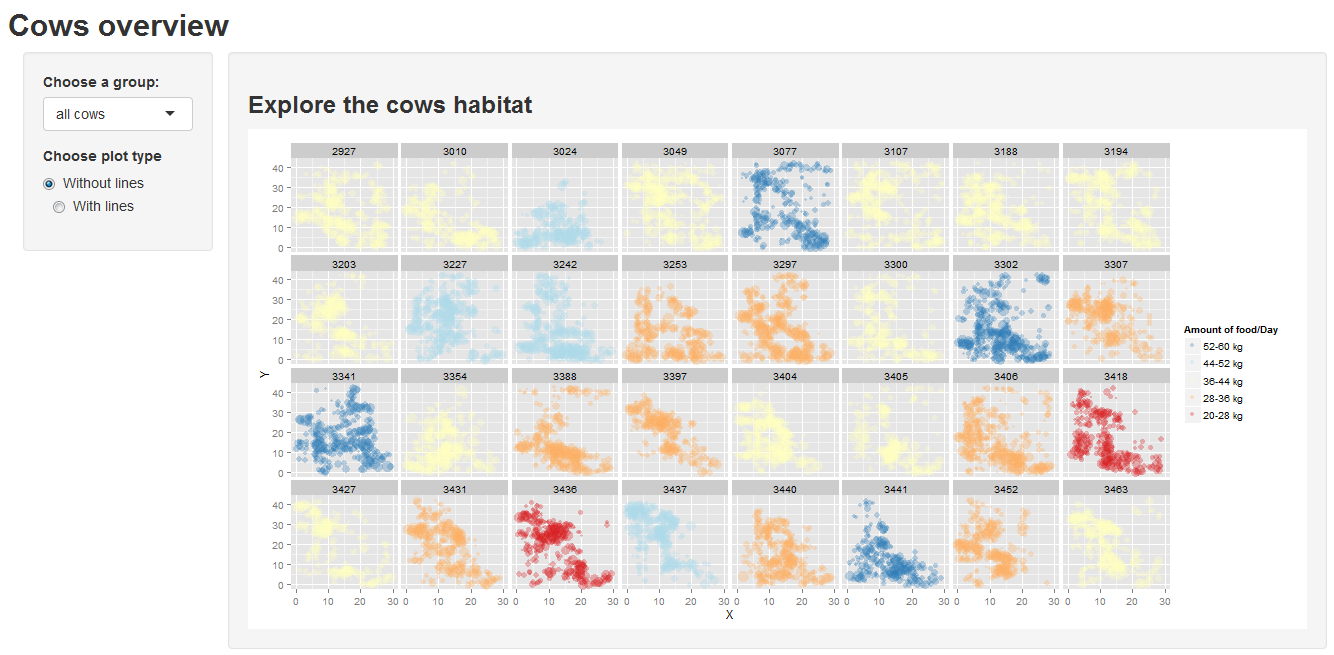
# מיפוי נתונים למימדים

להלן מיפוי הנתונים הסופי:

1. קואורדינאטות במרחב: מיקום. את הקואורדינאטות הדו ממדיות על ידי מיקום בתרשים דו מימדי, סקאלת הנתונים תהיה בעשרות מטרים בכדי לאפשר הבנה טובה יותר למתבונן (אין צורך במיקום ממש המדויק).
2. גודל: לוגריתם של משך השהייה בכל נקודה. גם כאן אין צורך במידע המדויק, אלא רק בכדי לזהות את האזורים העיקריים שבהם הפרה רובצת.
3. צבע: כמות המזון שפרה צורכת. ייעשה עי מיפוי לתת קבוצה מתוך חלוקה של 5 קבוצות כאשר 1- הפרה צורכת מעט מזון, 5- הפרה צורכת כמות גבוהה של מזון.
4. חלוקה לגרפים תיעשה על בסיס מספר הפרה - לכל פרה יוקצה גרף נפרד.
5. מידע משלים – השתמשנו בפקודה "geom\_density" בכדי לאפשר זיהוי קל יותר של אזורי הרביצה העיקריים בוויזואליזציה.

# תוצאות הוויזואליזציה

## גרפים

תצוגה של המסך הראשי:

תצוגה של המסך המשני:



## קוד

* קוד ggplot

library(reshape)

library(ggplot2)

Cp1<-read.csv("C:\\Users\\ilana\\Desktop\\vizILA\\allCows.csv")

LONG <- Cp1

LONG$catFood <- as.factor(LONG$catFood)

LONG$cow <- as.factor(LONG$cow)

# create grid of all cows

t <-ggplot(LONG, aes(x = LONG$x , y = LONG$y, color = catFood, size = LONG$logDuration )) + geom\_point(alpha = 0.3, position = "jitter", xlim=c(0,30)) + labs( x = "X" , y = "Y") + scale\_x\_continuous(limits=c(0, 30)) + facet\_wrap( ~ cow, ncol=8)+ scale\_fill\_continuous(guide = guide\_legend(title = "V")) + scale\_color\_brewer( name = "Amount of food/Day",labels = c("20-28 kg","28-36 kg", "36-44 kg", "44-52 kg","52-60 kg") , palette="RdYlBu")+ guides(size=FALSE, color = guide\_legend(reverse=TRUE) )

t

do1 <- LONG [ LONG$cow == 2927, ]

pCow1 <- ggplot(do1 , aes(x = do1$x , y = do1$y, size = do1$logDuration )) + geom\_point( alpha = 0.3, position = "jitter", xlim=c(0,30), colour="#2c7bb6") + labs( x = "X" , y = "Y") + scale\_x\_continuous(limits=c(0, 30)) + geom\_density2d(colour="#4d4d4d") + guides(size=guide\_legend(title= "Duration"))

do2 <- LONG [ LONG$cow == 3010, ]

pCow2 <- ggplot(do2 , aes(x = do2$x , y = do2$y, size = do2$logDuration )) + geom\_point( alpha = 0.3, position = "jitter", xlim=c(0,30),colour="#2c7bb6") + labs( x = "X" , y = "Y") + scale\_x\_continuous(limits=c(0, 30)) + geom\_density2d(colour="#4d4d4d") + guides(size=guide\_legend(title= "Duration"))

do3 <- LONG [ LONG$cow == 3024, ]

pCow3 <- ggplot(do3 , aes(x = do3$x , y = do3$y, size = do3$logDuration )) + geom\_point( alpha = 0.3, position = "jitter", xlim=c(0,30),colour="#2c7bb6") + labs( x = "X" , y = "Y") + scale\_x\_continuous(limits=c(0, 30)) + geom\_density2d(colour="#4d4d4d") + guides(size=guide\_legend(title= "Duration"))

do4 <- LONG [ LONG$cow == 3049, ]

pCow4 <- ggplot(do4 , aes(x = do4$x , y = do4$y, size = do4$logDuration )) + geom\_point( alpha = 0.3, position = "jitter", xlim=c(0,30),colour="#2c7bb6") + labs( x = "X" , y = "Y") + scale\_x\_continuous(limits=c(0, 30)) + geom\_density2d(colour="#4d4d4d") + guides(size=guide\_legend(title= "Duration"))

do5 <- LONG [ LONG$cow == 3077, ]

pCow5 <- ggplot(do5 , aes(x = do5$x , y = do5$y, size = do5$logDuration )) + geom\_point( alpha = 0.3, position = "jitter", xlim=c(0,30),colour="#2c7bb6") + labs( x = "X" , y = "Y") + scale\_x\_continuous(limits=c(0, 30)) + geom\_density2d(colour="#4d4d4d") + guides(size=guide\_legend(title= "Duration"))

do6 <- LONG [ LONG$cow == 3107, ]

pCow6 <- ggplot(do6 , aes(x = do6$x , y = do6$y, size = do6$logDuration )) + geom\_point( alpha = 0.3, position = "jitter", xlim=c(0,30),colour="#2c7bb6") + labs( x = "X" , y = "Y") + scale\_x\_continuous(limits=c(0, 30)) + geom\_density2d(colour="#4d4d4d") + guides(size=guide\_legend(title= "Duration"))

do7 <- LONG [ LONG$cow == 3188, ]

pCow7 <- ggplot(do7 , aes(x = do7$x , y = do7$y, size = do7$logDuration )) + geom\_point( alpha = 0.3, position = "jitter", xlim=c(0,30),colour="#2c7bb6") + labs( x = "X" , y = "Y") + scale\_x\_continuous(limits=c(0, 30)) + geom\_density2d(colour="#4d4d4d") + guides(size=guide\_legend(title= "Duration"))

do8 <- LONG [ LONG$cow == 3194, ]

pCow8 <- ggplot(do8 , aes(x = do8$x , y = do8$y, size = do8$logDuration )) + geom\_point( alpha = 0.3, position = "jitter", xlim=c(0,30),colour="#2c7bb6") + labs( x = "X" , y = "Y") + scale\_x\_continuous(limits=c(0, 30)) + geom\_density2d(colour="#4d4d4d") + guides(size=guide\_legend(title= "Duration"))

do9 <- LONG [ LONG$cow == 3203, ]

pCow9 <- ggplot(do9 , aes(x = do9$x , y = do9$y, size = do9$logDuration )) + geom\_point( alpha = 0.3, position = "jitter", xlim=c(0,30),colour="#2c7bb6") + labs( x = "X" , y = "Y") + scale\_x\_continuous(limits=c(0, 30)) + geom\_density2d(colour="#4d4d4d") + guides(size=guide\_legend(title= "Duration"))

do10 <- LONG [ LONG$cow == 3227, ]

pCow10 <- ggplot(do10 , aes(x = do10$x , y = do10$y, size = do10$logDuration )) + geom\_point( alpha = 0.3, position = "jitter", xlim=c(0,30),colour="#2c7bb6") + labs( x = "X" , y = "Y") + scale\_x\_continuous(limits=c(0, 30)) + geom\_density2d(colour="#4d4d4d") + guides(size=guide\_legend(title= "Duration"))

do11 <- LONG [ LONG$cow == 3242, ]

pCow11 <- ggplot(do11 , aes(x = do11$x , y = do11$y, size = do11$logDuration )) + geom\_point( alpha = 0.3, position = "jitter", xlim=c(0,30),colour="#2c7bb6") + labs( x = "X" , y = "Y") + scale\_x\_continuous(limits=c(0, 30)) + geom\_density2d(colour="#4d4d4d") + guides(size=guide\_legend(title= "Duration"))

do12 <- LONG [ LONG$cow == 3253, ]

pCow12 <- ggplot(do12 , aes(x = do12$x , y = do12$y, size = do12$logDuration )) + geom\_point( alpha = 0.3, position = "jitter", xlim=c(0,30),colour="#2c7bb6") + labs( x = "X" , y = "Y") + scale\_x\_continuous(limits=c(0, 30)) + geom\_density2d(colour="#4d4d4d") + guides(size=guide\_legend(title= "Duration"))

do13 <- LONG [ LONG$cow == 3297, ]

pCow13 <- ggplot(do13 , aes(x = do13$x , y = do13$y, size = do13$logDuration )) + geom\_point( alpha = 0.3, position = "jitter", xlim=c(0,30),colour="#2c7bb6") + labs( x = "X" , y = "Y") + scale\_x\_continuous(limits=c(0, 30)) + geom\_density2d(colour="#4d4d4d") + guides(size=guide\_legend(title= "Duration"))

do14 <- LONG [ LONG$cow == 3300, ]

pCow14 <- ggplot(do14 , aes(x = do14$x , y = do14$y, size = do14$logDuration )) + geom\_point( alpha = 0.3, position = "jitter", xlim=c(0,30),colour="#2c7bb6") + labs( x = "X" , y = "Y") + scale\_x\_continuous(limits=c(0, 30)) + geom\_density2d(colour="#4d4d4d") + guides(size=guide\_legend(title= "Duration"))

do15 <- LONG [ LONG$cow == 3302, ]

pCow15 <- ggplot(do15 , aes(x = do15$x , y = do15$y, size = do15$logDuration )) + geom\_point( alpha = 0.3, position = "jitter", xlim=c(0,30),colour="#2c7bb6") + labs( x = "X" , y = "Y") + scale\_x\_continuous(limits=c(0, 30)) + geom\_density2d(colour="#4d4d4d") + guides(size=guide\_legend(title= "Duration"))

do16 <- LONG [ LONG$cow == 3307, ]

pCow16 <- ggplot(do16 , aes(x = do16$x , y = do16$y, size = do16$logDuration )) + geom\_point( alpha = 0.3, position = "jitter", xlim=c(0,30),colour="#2c7bb6") + labs( x = "X" , y = "Y") + scale\_x\_continuous(limits=c(0, 30)) + geom\_density2d(colour="#4d4d4d") + guides(size=guide\_legend(title= "Duration"))

do17 <- LONG [ LONG$cow == 3341, ]

pCow17 <- ggplot(do17 , aes(x = do17$x , y = do17$y, size = do17$logDuration )) + geom\_point( alpha = 0.3, position = "jitter", xlim=c(0,30),colour="#2c7bb6") + labs( x = "X" , y = "Y") + scale\_x\_continuous(limits=c(0, 30)) + geom\_density2d(colour="#4d4d4d") + guides(size=guide\_legend(title= "Duration"))

do18 <- LONG [ LONG$cow == 3354, ]

pCow18 <- ggplot(do18 , aes(x = do18$x , y = do18$y, size = do18$logDuration )) + geom\_point( alpha = 0.3, position = "jitter", xlim=c(0,30),colour="#2c7bb6") + labs( x = "X" , y = "Y") + scale\_x\_continuous(limits=c(0, 30)) + geom\_density2d(colour="#4d4d4d") + guides(size=guide\_legend(title= "Duration"))

do19 <- LONG [ LONG$cow == 3388, ]

pCow19 <- ggplot(do19 , aes(x = do19$x , y = do19$y, size = do19$logDuration )) + geom\_point( alpha = 0.3, position = "jitter", xlim=c(0,30),colour="#2c7bb6") + labs( x = "X" , y = "Y") + scale\_x\_continuous(limits=c(0, 30)) + geom\_density2d(colour="#4d4d4d") + guides(size=guide\_legend(title= "Duration"))

do20 <- LONG [ LONG$cow == 3397, ]

pCow20 <- ggplot(do20 , aes(x = do20$x , y = do20$y, size = do20$logDuration )) + geom\_point( alpha = 0.3, position = "jitter", xlim=c(0,30),colour="#2c7bb6") + labs( x = "X" , y = "Y") + scale\_x\_continuous(limits=c(0, 30)) + geom\_density2d(colour="#4d4d4d") + guides(size=guide\_legend(title= "Duration"))

do21 <- LONG [ LONG$cow == 3404, ]

pCow21 <- ggplot(do21 , aes(x = do21$x , y = do21$y, size = do21$logDuration )) + geom\_point( alpha = 0.3, position = "jitter", xlim=c(0,30),colour="#2c7bb6") + labs( x = "X" , y = "Y") + scale\_x\_continuous(limits=c(0, 30)) + geom\_density2d(colour="#4d4d4d") + guides(size=guide\_legend(title= "Duration"))

do22 <- LONG [ LONG$cow == 3405, ]

pCow22 <- ggplot(do22 , aes(x = do22$x , y = do22$y, size = do22$logDuration )) + geom\_point( alpha = 0.3, position = "jitter", xlim=c(0,30),colour="#2c7bb6") + labs( x = "X" , y = "Y") + scale\_x\_continuous(limits=c(0, 30)) + geom\_density2d(colour="#4d4d4d") + guides(size=guide\_legend(title= "Duration"))

do23 <- LONG [ LONG$cow == 3406, ]

pCow23 <- ggplot(do23 , aes(x = do23$x , y = do23$y, size = do23$logDuration )) + geom\_point( alpha = 0.3, position = "jitter", xlim=c(0,30),colour="#2c7bb6") + labs( x = "X" , y = "Y") + scale\_x\_continuous(limits=c(0, 30)) + geom\_density2d(colour="#4d4d4d") + guides(size=guide\_legend(title= "Duration"))

do24 <- LONG [ LONG$cow == 3418, ]

pCow24 <- ggplot(do24 , aes(x = do24$x , y = do24$y, size = do24$logDuration )) + geom\_point( alpha = 0.3, position = "jitter", xlim=c(0,30),colour="#2c7bb6") + labs( x = "X" , y = "Y") + scale\_x\_continuous(limits=c(0, 30)) + geom\_density2d(colour="#4d4d4d") + guides(size=guide\_legend(title= "Duration"))

do25 <- LONG [ LONG$cow == 3427, ]

pCow25 <- ggplot(do25 , aes(x = do25$x , y = do25$y, size = do25$logDuration )) + geom\_point( alpha = 0.3, position = "jitter", xlim=c(0,30),colour="#2c7bb6") + labs( x = "X" , y = "Y") + scale\_x\_continuous(limits=c(0, 30)) + geom\_density2d(colour="#4d4d4d") + guides(size=guide\_legend(title= "Duration"))

do26 <- LONG [ LONG$cow == 3431, ]

pCow26 <- ggplot(do26 , aes(x = do26$x , y = do26$y, size = do26$logDuration )) + geom\_point( alpha = 0.3, position = "jitter", xlim=c(0,30),colour="#2c7bb6") + labs( x = "X" , y = "Y") + scale\_x\_continuous(limits=c(0, 30)) + geom\_density2d(colour="#4d4d4d") + guides(size=guide\_legend(title= "Duration"))

do27 <- LONG [ LONG$cow == 3436, ]

pCow27 <- ggplot(do27 , aes(x = do27$x , y = do27$y, size = do27$logDuration )) + geom\_point( alpha = 0.3, position = "jitter", xlim=c(0,30),colour="#2c7bb6") + labs( x = "X" , y = "Y") + scale\_x\_continuous(limits=c(0, 30)) + geom\_density2d(colour="#4d4d4d") + guides(size=guide\_legend(title= "Duration"))

do28 <- LONG [ LONG$cow == 3437, ]

pCow28 <- ggplot(do28 , aes(x = do28$x , y = do28$y, size = do28$logDuration )) + geom\_point( alpha = 0.3, position = "jitter", xlim=c(0,30),colour="#2c7bb6") + labs( x = "X" , y = "Y") + scale\_x\_continuous(limits=c(0, 30)) + geom\_density2d(colour="#4d4d4d") + guides(size=guide\_legend(title= "Duration"))

do29 <- LONG [ LONG$cow == 3440, ]

pCow29 <- ggplot(do29 , aes(x = do29$x , y = do29$y, size = do29$logDuration )) + geom\_point( alpha = 0.3, position = "jitter", xlim=c(0,30),colour="#2c7bb6") + labs( x = "X" , y = "Y") + scale\_x\_continuous(limits=c(0, 30)) + geom\_density2d(colour="#4d4d4d") + guides(size=guide\_legend(title= "Duration"))

do30 <- LONG [ LONG$cow == 3441, ]

pCow30 <- ggplot(do30 , aes(x = do30$x , y = do30$y, size = do30$logDuration )) + geom\_point( alpha = 0.3, position = "jitter", xlim=c(0,30),colour="#2c7bb6") + labs( x = "X" , y = "Y") + scale\_x\_continuous(limits=c(0, 30)) + geom\_density2d(colour="#4d4d4d") + guides(size=guide\_legend(title= "Duration"))

do31 <- LONG [ LONG$cow == 3452, ]

pCow31 <- ggplot(do31 , aes(x = do31$x , y = do31$y, size = do31$logDuration )) + geom\_point( alpha = 0.3, position = "jitter", xlim=c(0,30),colour="#2c7bb6") + labs( x = "X" , y = "Y") + scale\_x\_continuous(limits=c(0, 30)) + geom\_density2d(colour="#4d4d4d") + guides(size=guide\_legend(title= "Duration"))

do32 <- LONG [ LONG$cow == 3463, ]

pCow32 <- ggplot(do32 , aes(x = do32$x , y = do32$y, size = do32$logDuration )) + geom\_point( alpha = 0.3, position = "jitter", xlim=c(0,30),colour="#2c7bb6") + labs( x = "X" , y = "Y") + scale\_x\_continuous(limits=c(0, 30)) + geom\_density2d(colour="#4d4d4d") + guides(size=guide\_legend(title= "Duration"))

#################################

fat1 <- LONG [ LONG$catFood == 1, ]

Pfat1 <- ggplot(fat1, aes(x = fat1$x , y = fat1$y , size = fat1$logDuration )) + geom\_point( colour = '#d7191c',alpha = 0.3, position = "jitter", xlim=c(0,30)) + labs( x = "X" , y = "Y") + scale\_x\_continuous(limits=c(0, 30)) + facet\_wrap( ~ cow, ncol=2) + guides(size=guide\_legend(title= "Duration"))

fat2 <- LONG [ LONG$catFood == 2, ]

Pfat2 <- ggplot(fat2, aes(x = fat2$x , y = fat2$y, size = fat2$logDuration )) + geom\_point( colour = '#fdae61',alpha = 0.3, position = "jitter", xlim=c(0,30)) + labs( x = "X" , y = "Y") + scale\_x\_continuous(limits=c(0, 30)) + facet\_wrap( ~ cow, ncol=3) + guides(size=guide\_legend(title= "Duration"))

fat3 <- LONG [ LONG$catFood == 3, ]

Pfat3 <- ggplot(fat3, aes(x = fat3$x , y = fat3$y, size = fat3$logDuration )) + geom\_point( colour = '#ffffbf', alpha = 0.3, position = "jitter", xlim=c(0,30)) + labs( x = "X" , y = "Y") + scale\_x\_continuous(limits=c(0, 30)) + facet\_wrap( ~ cow, ncol=4) + guides(size=guide\_legend(title= "Duration"))

fat4 <- LONG [ LONG$catFood == 4, ]

Pfat4 <- ggplot(fat4, aes(x = fat4$x , y = fat4$y, size = fat4$logDuration )) + geom\_point( colour = '#abd9e9', alpha = 0.3, position = "jitter", xlim=c(0,30)) + labs( x = "X" , y = "Y") + scale\_x\_continuous(limits=c(0, 30)) + facet\_wrap( ~ cow, ncol=2) + guides(size=guide\_legend(title= "Duration"))

fat5 <- LONG [ LONG$catFood == 5, ]

Pfat5 <- ggplot(fat5, aes(x = fat5$x , y = fat5$y, size = fat5$logDuration )) + geom\_point( colour = '#2c7bb6',alpha = 0.3, position = "jitter", xlim=c(0,30)) + labs( x = "X" , y = "Y") + scale\_x\_continuous(limits=c(0, 30)) + facet\_wrap( ~ cow, ncol=2) + guides(size=guide\_legend(title= "Duration"))

* קוד shiny

library(shiny)

library(ggplot2)

#ui

ui <- fluidPage(titlePanel(strong("Cows overview")),sidebarPanel(width= 2,

selectInput("dataset", "Choose a group:",

choices = c("all cows", "group1", "group2", "group3" , "group4", "group5")),

radioButtons("plot\_type", "Choose plot type",

c("Without lines", "With lines"), inline = TRUE)

),

#fluidRow(

column(width = 10, class = "well",

h3(strong("Explore the cows habitat")),

plotOutput("plot1", height = 500

)

),

# ),

column(width = 12, class = "well",

selectInput("cow", "Choose a specific cow:",

choices = c("cow2927", "cow3010", "cow3024", "cow3049" , "cow3077", "cow3107", "cow3188",

"cow3194", "cow3203", "cow3227", "cow3242" , "cow3253", "cow3297", "cow3300",

"cow3302", "cow3307", "cow3341", "cow3354" , "cow3388", "cow3397", "cow3404",

"cow3405", "cow3406", "cow3418", "cow3427" , "cow3431", "cow3436", "cow3437",

"cow3440", "cow3441", "cow3452", "cow3463")),

h3(strong("Expore one cow- left plot controls right plot")),

fluidRow(

column(width = 4,

plotOutput("plot2", height = 300,

brush = brushOpts(

id = "plot2\_brush",

resetOnNew = TRUE

)

)

),

column(width = 4,

plotOutput("plot3", height = 300)

)

)

)

)

#------------------------------------------------------------

#server

server <- function(input, output) {

# Single zoomable plot (on left)

ranges <- reactiveValues(x = NULL, y = NULL)

output$plot1 <- renderPlot({

if (input$dataset== "all cows"){

if (input$plot\_type == "Without lines") {

t

} else if (input$plot\_type == "With lines") {

t + geom\_density2d(colour=1)

}

}else if (input$dataset =="group1"){

if (input$plot\_type == "Without lines") {

Pfat1

} else if (input$plot\_type == "With lines") {

Pfat1 + geom\_density2d(colour=1)

}

}else if (input$dataset =="group2"){

if (input$plot\_type == "Without lines") {

Pfat2

} else if (input$plot\_type == "With lines") {

Pfat2 + geom\_density2d(colour=1)

}

}else if (input$dataset =="group3"){

if (input$plot\_type == "Without lines") {

Pfat3

} else if (input$plot\_type == "With lines") {

Pfat3 + geom\_density2d(colour=1)

}

}else if (input$dataset =="group4"){

if (input$plot\_type == "Without lines") {

Pfat4

} else if (input$plot\_type == "With lines") {

Pfat4 + geom\_density2d(colour=1)

}

}else if (input$dataset =="group5"){

if(input$plot\_type=="Without lines"){

Pfat5

}else if (input$plot\_type == "With lines") {

Pfat5 + geom\_density2d(colour=1)

}

}

})

# -------------------------------------------------------------------

# Linked plots (middle and right)

ranges2 <- reactiveValues(x = NULL, y = NULL)

output$plot2 <- renderPlot({ if (input$cow== "cow2927"){pCow1}

else if (input$cow== "cow3010"){pCow2}

else if (input$cow== "cow3024"){pCow3}

else if (input$cow== "cow3049"){pCow4}

else if (input$cow== "cow3077"){pCow5}

else if (input$cow== "cow3107"){pCow6}

else if (input$cow== "cow3188"){pCow7}

else if (input$cow== "cow3194"){pCow8}

else if (input$cow== "cow3203"){pCow9}

else if (input$cow== "cow3227"){pCow10}

else if (input$cow== "cow3242"){pCow11}

else if (input$cow== "cow3253"){pCow12}

else if (input$cow== "cow3297"){pCow13}

else if (input$cow== "cow3300"){pCow14}

else if (input$cow== "cow3302"){pCow15}

else if (input$cow== "cow3307"){pCow16}

else if (input$cow== "cow3341"){pCow17}

else if (input$cow== "cow3354"){pCow18}

else if (input$cow== "cow3388"){pCow19}

else if (input$cow== "cow3397"){pCow20}

else if (input$cow== "cow3404"){pCow21}

else if (input$cow== "cow3405"){pCow22}

else if (input$cow== "cow3406"){pCow23}

else if (input$cow== "cow3418"){pCow24}

else if (input$cow== "cow3427"){pCow25}

else if (input$cow== "cow3431"){pCow26}

else if (input$cow== "cow3436"){pCow27}

else if (input$cow== "cow3437"){pCow28}

else if (input$cow== "cow3440"){pCow29}

else if (input$cow== "cow3441"){pCow30}

else if (input$cow== "cow3452"){pCow31}

else if (input$cow== "cow3463"){pCow32}

})

output$plot3 <- renderPlot({if (input$cow== "cow2927"){pCow1+ coord\_cartesian(xlim = ranges2$x, ylim = ranges2$y)}

else if (input$cow=="cow3010"){pCow2+ coord\_cartesian(xlim = ranges2$x, ylim = ranges2$y)}

else if (input$cow=="cow3024"){pCow3+ coord\_cartesian(xlim = ranges2$x, ylim = ranges2$y)}

else if (input$cow=="cow3049"){pCow4+ coord\_cartesian(xlim = ranges2$x, ylim = ranges2$y)}

else if (input$cow=="cow3077"){pCow5+ coord\_cartesian(xlim = ranges2$x, ylim = ranges2$y)}

else if (input$cow=="cow3107"){pCow6+ coord\_cartesian(xlim = ranges2$x, ylim = ranges2$y)}

else if (input$cow=="cow3188"){pCow7+ coord\_cartesian(xlim = ranges2$x, ylim = ranges2$y)}

else if (input$cow== "cow3194"){pCow8+ coord\_cartesian(xlim = ranges2$x, ylim = ranges2$y)}

else if (input$cow== "cow3203"){pCow9+ coord\_cartesian(xlim = ranges2$x, ylim = ranges2$y)}

else if (input$cow== "cow3227"){pCow10+ coord\_cartesian(xlim = ranges2$x, ylim = ranges2$y)}

else if (input$cow== "cow3242"){pCow11+ coord\_cartesian(xlim = ranges2$x, ylim = ranges2$y)}

else if (input$cow== "cow3253"){pCow12+ coord\_cartesian(xlim = ranges2$x, ylim = ranges2$y)}

else if (input$cow== "cow3297"){pCow13+ coord\_cartesian(xlim = ranges2$x, ylim = ranges2$y)}

else if (input$cow== "cow3300"){pCow14+ coord\_cartesian(xlim = ranges2$x, ylim = ranges2$y)}

else if (input$cow== "cow3302"){pCow15+ coord\_cartesian(xlim = ranges2$x, ylim = ranges2$y)}

else if (input$cow== "cow3307"){pCow16+ coord\_cartesian(xlim = ranges2$x, ylim = ranges2$y)}

else if (input$cow== "cow3341"){pCow17+ coord\_cartesian(xlim = ranges2$x, ylim = ranges2$y)}

else if (input$cow== "cow3354"){pCow18+ coord\_cartesian(xlim = ranges2$x, ylim = ranges2$y)}

else if (input$cow== "cow3388"){pCow19+ coord\_cartesian(xlim = ranges2$x, ylim = ranges2$y)}

else if (input$cow== "cow3397"){pCow20+ coord\_cartesian(xlim = ranges2$x, ylim = ranges2$y)}

else if (input$cow== "cow3404"){pCow21+ coord\_cartesian(xlim = ranges2$x, ylim = ranges2$y)}

else if (input$cow== "cow3405"){pCow22+ coord\_cartesian(xlim = ranges2$x, ylim = ranges2$y)}

else if (input$cow== "cow3406"){pCow23+ coord\_cartesian(xlim = ranges2$x, ylim = ranges2$y)}

else if (input$cow== "cow3418"){pCow24+ coord\_cartesian(xlim = ranges2$x, ylim = ranges2$y)}

else if (input$cow== "cow3427"){pCow25+ coord\_cartesian(xlim = ranges2$x, ylim = ranges2$y)}

else if (input$cow== "cow3431"){pCow26+ coord\_cartesian(xlim = ranges2$x, ylim = ranges2$y)}

else if (input$cow== "cow3436"){pCow27+ coord\_cartesian(xlim = ranges2$x, ylim = ranges2$y)}

else if (input$cow== "cow3437"){pCow28+ coord\_cartesian(xlim = ranges2$x, ylim = ranges2$y)}

else if (input$cow== "cow3440"){pCow29+ coord\_cartesian(xlim = ranges2$x, ylim = ranges2$y)}

else if (input$cow== "cow3441"){pCow30+ coord\_cartesian(xlim = ranges2$x, ylim = ranges2$y)}

else if (input$cow== "cow3452"){pCow31+ coord\_cartesian(xlim = ranges2$x, ylim = ranges2$y)}

else if (input$cow== "cow3463"){pCow32+ coord\_cartesian(xlim = ranges2$x, ylim = ranges2$y)}

})

# When a double-click happens, check if there's a brush on the plot.

# If so, zoom to the brush bounds; if not, reset the zoom.

observe({

brush <- input$plot2\_brush

if (!is.null(brush)) {

ranges2$x <- c(brush$xmin, brush$xmax)

ranges2$y <- c(brush$ymin, brush$ymax)

} else {

ranges2$x <- NULL

ranges2$y <- NULL

}

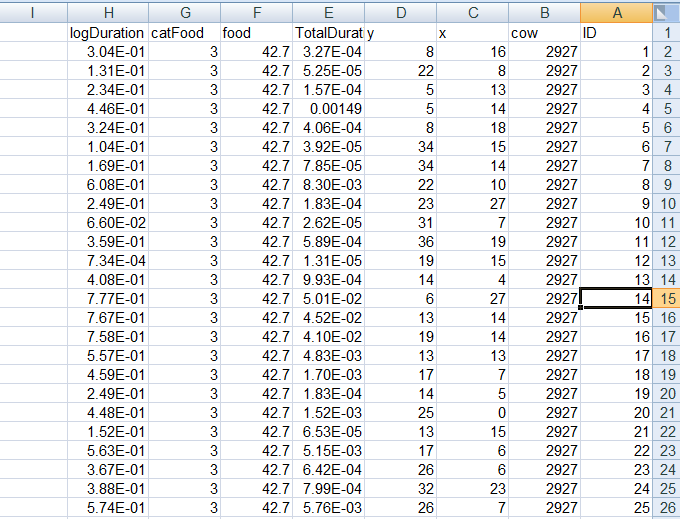
})#observe

}#function (input,output)

shinyApp(ui, server)

## קבצי עזר

טבלה חלקית של נתוני הפרות הסופיים:



# סיכום ומסקנות

## ערך הוויזואליזציה

נבחן את ערכה של הוויזואליזציה לפי המדדים אשר נלמדו בכיתה:

* יכולת לחסוך בזמן – מכיוון שוויזואליזציה זו מאגדת בתוכה נתונים רבים, אשר קשה מאוד לקרוא מתוך טבלה ולהסיק על בסיסם מסקנות, אין ספק שהוויזואליזציה חוסכת זמן במענה על שאלת הוויזואליזציה.
* יכולת להבין תובנות חדשות – כאשר ניגשנו לוויזואליזציה המטרה הייתה למצוא קשר בין כמות המזון שפרה צורכת לבין ההתנהגות המרחבית שלה, מתוך הוויזואליזציה ראינו שלא ניתן לראות קשר ישיר שכזה באופן מובהק ובולט לעין, והרי גם אי מציאת קשר זוהי תובנה, כמו כן מבנה הוויזואליזציה אשר מציג כל פרה בנפרד מאפשר גם הסקת מסקנות חדשות, ספציפיות לכל פרה וגם זיהוי חריגות מהכלל.
* מהות, סיכום כל הידע והעברתו באופן תמציתי – הוויזואליזציה אשר יצרנו אכן מכילה תמצית של עיקר הידע ולמעשה מאגדת בתוכה כמות גדולה מאוד של מידע, אשר קשה היה לחקור בצורה אחרת, עם זאת ישנו מידע אשר היה בידנו ובחרנו שלא לעשות בו שימוש (כדוגמת המידע האם הפרה דורשת או חלוקה למקטעים ביום), בסופו של דבר לא נמצא קשר ויכול להיות שחקר המידע ברזולוציות או בהקשרים נוספים כן היה מביא לתוצאה אחרת.
* אמון בתוצאות הוויזואליזציה – הוויזואליזציה מציגה 32 פרות שונות אשר בתוכן חלוקה לתת קבוצות לפי כמות המזון שהן צורכות, הוויזואליזציה מספקת מושג לגבי התנהגותן המרחבית של הפרות ולא מראה קשר בין השתייכות לקבוצה לבין ההתנהגות המרחבית אך היא אינה מכילה כמות מספקת של נתונים כדי ליצור אמון מספק בתוצאותיה ולהניח כי אלו חד משמעיות. ניתן לומר שהוויזואליזציה מספקת תחושה ומאפשרת לחוקר לזהות כיווני מחקר שונים וחלוקות שונות, אך אין כאן מובהקות לגבי תוצאה כזו או אחרת שהחוקר יכול ממש להסתמך עליה.

## יתרונות

* היתרון העיקרי של הוויזואליזציה היא ביכולתה לאגד מידע רב ולהציגו בתצוגה אחת, תצוגה זו חוסכת לחוקר זמן רב ומאפשרת זיהוי מגמות וחריגים באופן פשוט ונוח.
* הוויזואליזציה כפי שיוצרה בעזרת הממשק האינטראקטיבי Shiny מאפשרת לחוקר חופש פעולה ומרווח תמרון בבחירת המידע שברצונו לחקור, לעניות דעתנו מדובר ביתרון אשר לוקח בחשבון את הצופה בוויזואליזציה, יכולותיו וציפיותיו מהוויזואליזציה.
* הבחנה בין עיקר לתפל – הוויזואליזציה עוצבה תוך הסתכלות מתמדת על שאלת המחקר, כך שעבור המשתנים העיקריים הושם דגש - מיקום ביחס לכמות המזון שהפרה צורכת, החלוקה לתרשימים נפרדים לכל פרה מבליטה את המשתנים הללו, כמו כן לא השתמשנו במידע אשר לא תרם לשאלת הוויזואליזציה מה שהשאיר את המשתנים הנחקרים בולטים וברורים.
* סקלאביליות – מכיוון שהוויזואליזציה יוצרה ב- Shiny מתאפשרות דרגות חופש שונות - ניתן להציג תת קבוצה של תרשימים או לחלופין לתחקר תרשים של פרה אחת בלבד, כמו כן ניתן להתאים וויזואליזציה זו לכמות משתנה של פרות\קבוצות.

## חסרונות

* חוסר יכולת לייצר תשובה מובהקת – מכיוון שביצענו חלוקה לקבוצות וחלוקה נוספת לפי כל פרה קשה מאוד להסיק מסקנות חד משמעיות מהוויזואליזציה, והיא למעשה נשארת רק בגדר מידע אשר מרמז על מגמות וחלוקות כלשהן, אך לא יכול להצביע עליהן בבירור. חיסרון זה נובע מהגרעיניות הגבוהה שבה המידע מוצג - כל פרה מוצגת בנפרד, ולכן קשה להסיק מסקנות על קבוצה.
* דיוק – על מנת לייצר וויזואליזציה קריאה וברורה הימרנו את משך השהייה בכל קואורדינאטה על ידי שימוש בלוגריתם, על ידי כך נוצר קושי בזיהוי ההבדלים במיפוי לגודל - בעיקר כאשר הגרפים קטנים.