Question 1: Space filling curves for NN-queries (1 P.)

a) Implement space-filling z-curves in a language of your choice. Your program has to take two files which contain points as input. The first file contains all the points of the base data set, while the second file contains query points. The program has to calculate and display:

- The k-NN, of each query point, in the base data set, based on the actual distance.
- The k-NN, of each query point, in the base data set, based on the z-curve distance.

You can use the template in OLAT, which already parses the files and provides utility classes. Submit the code and the output of your program when executed with $k = 3$ and the two data files provided in OLAT. If you do not use the template, also submit instructions on how to compile and execute your program.

Note: If you use code from external sources, provide the source as a comment.

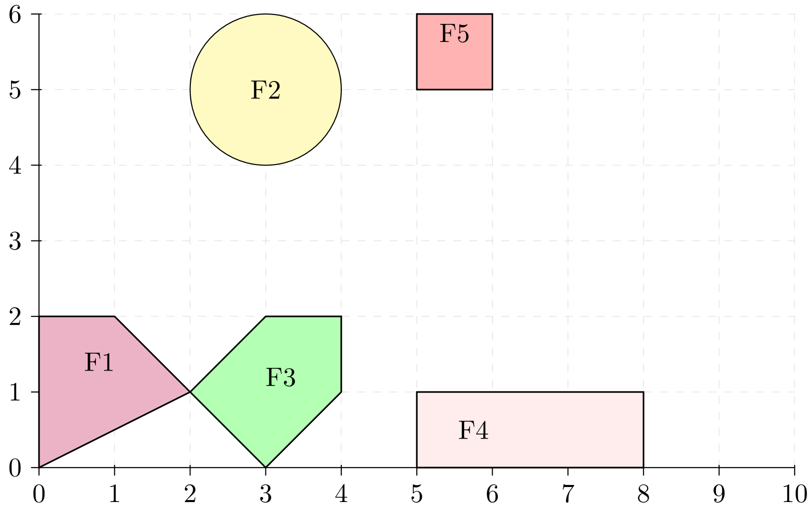
Required submission: Source Code in separate file; Output after executing code with provided data files; (Compile/Execute instructions if not using template).

b) Which differences can you see between the two results of your implementation? Explain why or why not the results are the same.

Required submission: Explanation

Question 2: R Tree (1 P.)

Given the following two-dimensional objects:



For the following R tree operations, explain exactly which steps are performed:

a) *Store the objects F1, F2, F3, F4, F5 in an initially empty R tree.* One node fits 1–2 entries.

Required submission: Final tree; Explanation of every insertion step;

b) *Find all objects, containing the point (3,1).*

Required submission: Objects that contain the point; Explanation of the search steps

c) *Find all objects that are positioned completely in the rectangle Q, which is defined by the points* $(3,4)(7,7)$.

Required submission: Objects that are positioned in Q; Explanation of search steps;

d) *Find all objects, intersecting with Q':* $(1,0)(3,6)$.

Required submission: Objects intersecting with Q'; Explanation of search step;

We assume that sharing exactly one point, also counts as intersecting.

Question 3: Index Structures in Metric Space (1 P.)

Show, for each of the following distance functions, that the properties for being a metric are fulfilled, or provide a counter example.

Required submission: Proof or counter example;

a) Manhattan/Taxicab distance for two vectors (or points) $a = (x_a, y_a)$ and $b = (x_b, y_b)$ calculated as $d(a, b) \mapsto |x_a - x_b| + |y_a - y_b|$.

b) The difference of set lengths: $d_L(s_1, s_2) \mapsto (|S(s_1)| - |S(s_2)|)^3$
With S being the set of characters of the string (e.g.: $S(\text{"Codd"}) = \{\text{'C', 'o', 'd'}\}$).

c) Sørensen–Dice coefficient, using the set of characters of the string:
$d_J(s_1, s_2) \mapsto |S(s_1) \cap S(s_2)| \,/\, |S(s_1)| + |S(s_2)|$.

Question 4: Misc Metric Indexing (1 P.)

Below are the points used for the following questions:

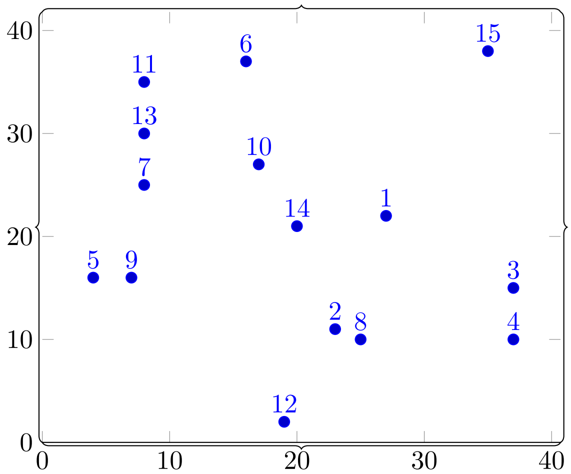
$$1:(27,22) \quad 2:(23,11) \quad 3:(37,15) \quad 4:(37,10) \quad 5:(4,16)$$

$$6:(16,37) \quad 7:(8,25) \quad 8:(25,10) \quad 9:(7,16) \quad 10:(17,27)$$

$$11:(8,35) \quad 12:(19,2) \quad 13:(8,30) \quad 14:(20,21) \quad 15:(35,38)$$

a) **GH Tree partitioning:** Create a GH partitioning, such that the leaf nodes of the tree have at most 2 elements. Draw the tree and draw the partitioning into the plot below.

Required submission: GH Tree; Partitioned plot;



b) **VP Tree:** Given the following VP tree, search for query point (25,4) with $\varepsilon = 11$. Describe which parts of the tree you pruned and why.

Required submission: Result of search; Explanation of search and pruning steps;