

Lab 2

Antonio Jurlina

09/19/2020

Data Overview

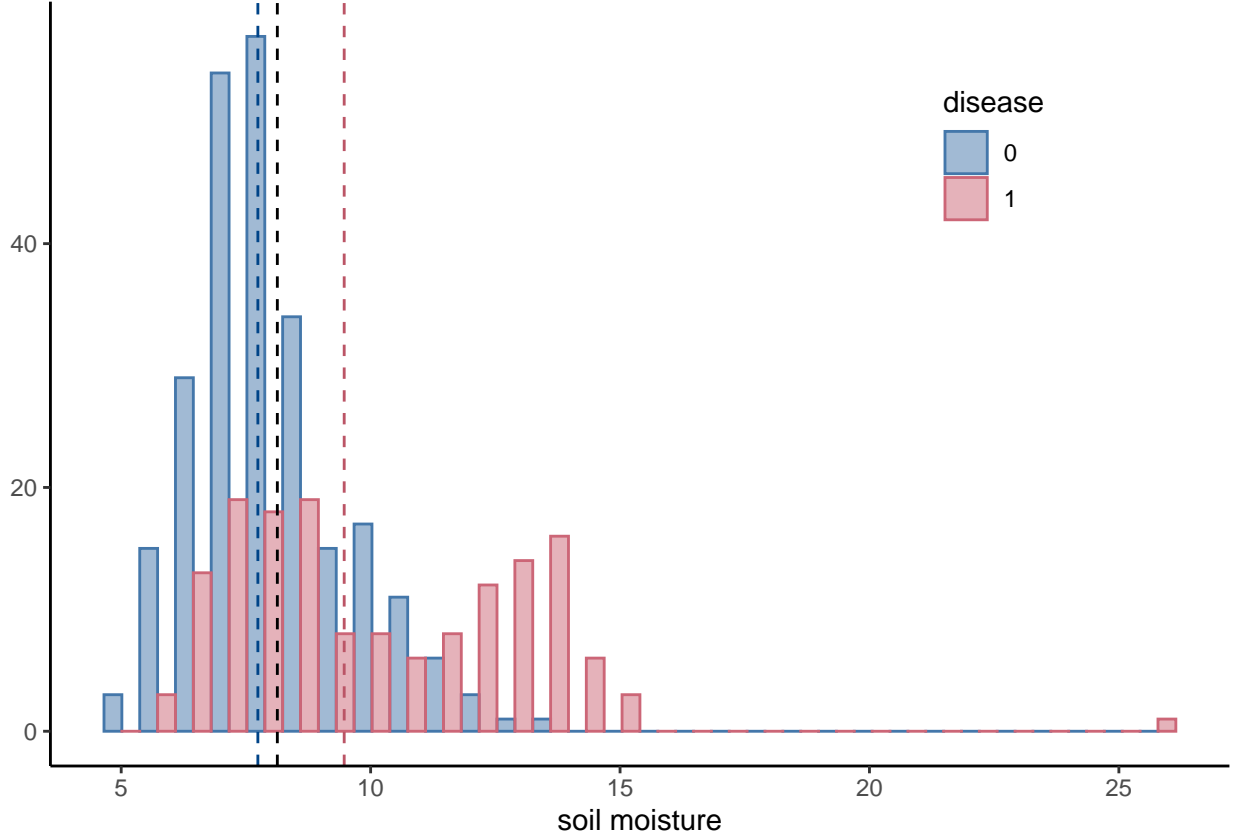


Figure 1: soil moisture distribution (with medians)

Table 1: summary statistics

data	min	median	mean	max	variance	skewness	kurtosis
complete	5.23	8.13	8.805412	26.00	5.682385	1.6310977	5.9559019
disease	5.82	9.47	10.066818	26.00	8.496823	1.1483211	3.8461556
no disease	5.23	7.74	8.015752	13.26	2.321733	0.8017359	0.4955825

Plot in Figure 1 is a histogram of all soil moisture data points, colored by whether there is known phytophthora presence in that area of measurement. Dashed lines represent the medians. The black one is the median for the entire data set, while the other two represent the medians of groups with and without phytophthora. As can be seen, there is a significant outlier with soil moisture, with phytophthora present. All relevant descriptive values can be found in Table 1.

Figure 2 represents the bird's eye view of the data distribution, with size and fill of each circle corresponding to the moisture level, and the edges of circles corresponding to the presence of disease. As can be seen, most of the moisture can be found in one corner of the field (with phytophthora present there as well). The outlier (where soil moisture is highest) can be found in

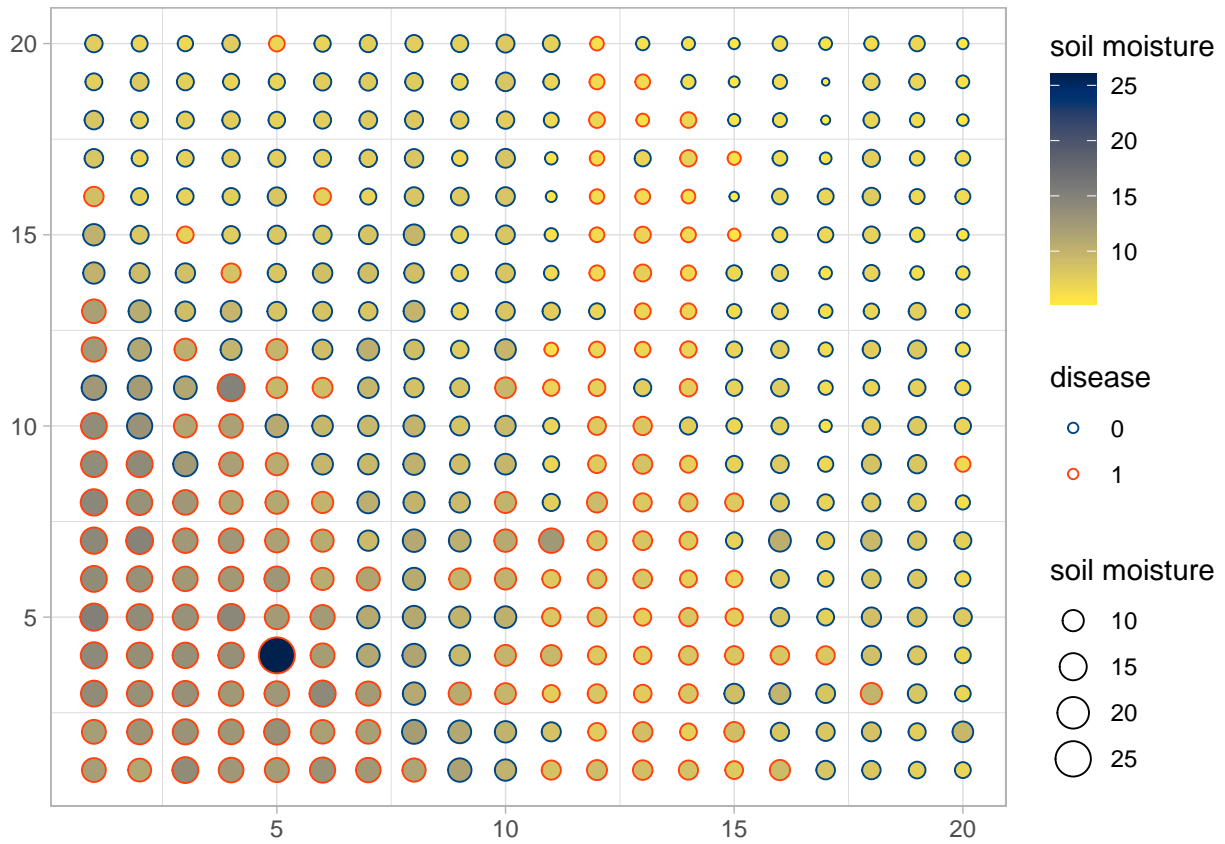


Figure 2: spatial patterns

this very corner. Figure 3 is the same data, just plotted in 3rd space, with soil moisture (left) and phytophthora (right). The black spatial patterns in the right graph indicate the presence of phytophthora and they parallel the presence of high soil moisture levels fairly well.

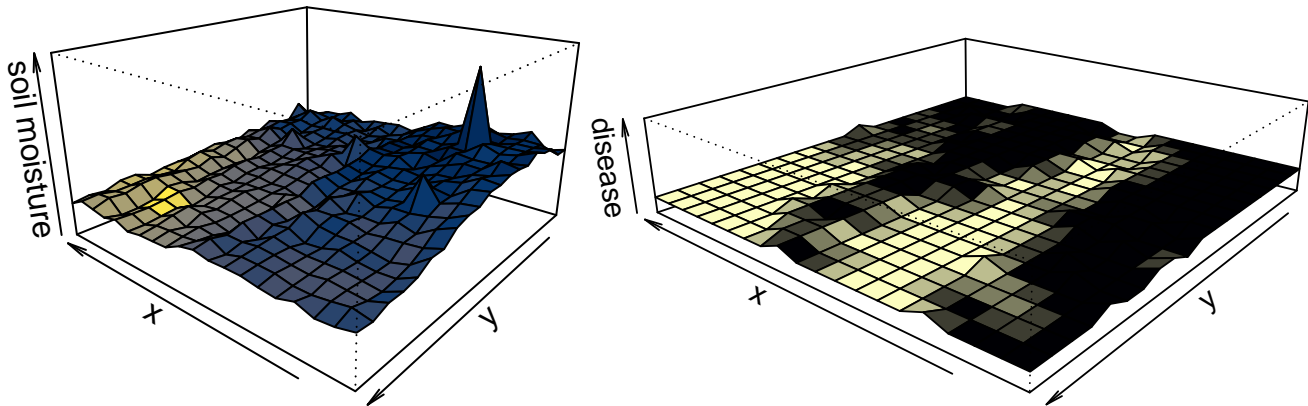


Figure 3: soil moisture data (left) vs. disease presence data (right)

OLS

Table 2: OLS estimates (dep variable = soil moisture)

term	estimate	std.error	statistic	p.value
(Intercept)	13.8819178	0.1844762	75.25046	0
X	-0.2550306	0.0115808	-22.02191	0
Y	-0.2284461	0.0115808	-19.72633	0

r.squared	adj.r.squared
0.6876704	0.6860969

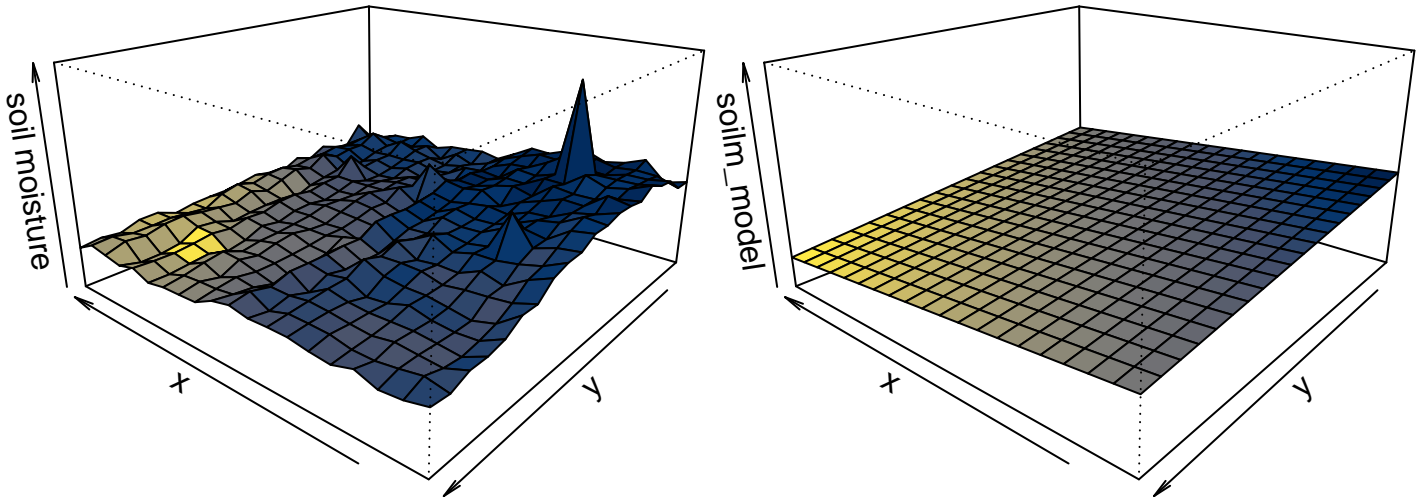


Figure 4: data (left) vs. OLS fit (right)

Looking at the linear regression results of a model in which soil moisture is the dependent variable (Table 2, Figure 4), we can conclude the model does a reasonable job of predicting soil moisture (and phytophthora by extension) presence. However, the averaging nature of this model fails to handle the outlier efficiently and therefore adds too much of an angle to the data plane indicating that almost all soil moisture is concentrated within that one corner. In reality, it is much more spread out. According to the OLS model, we can expect to see around 0.26 and 0.23 unit increases in moisture (on average) for each unit of space we get closer to the corner with the outlier in it, across X and Y axes respectively.

Median Polish

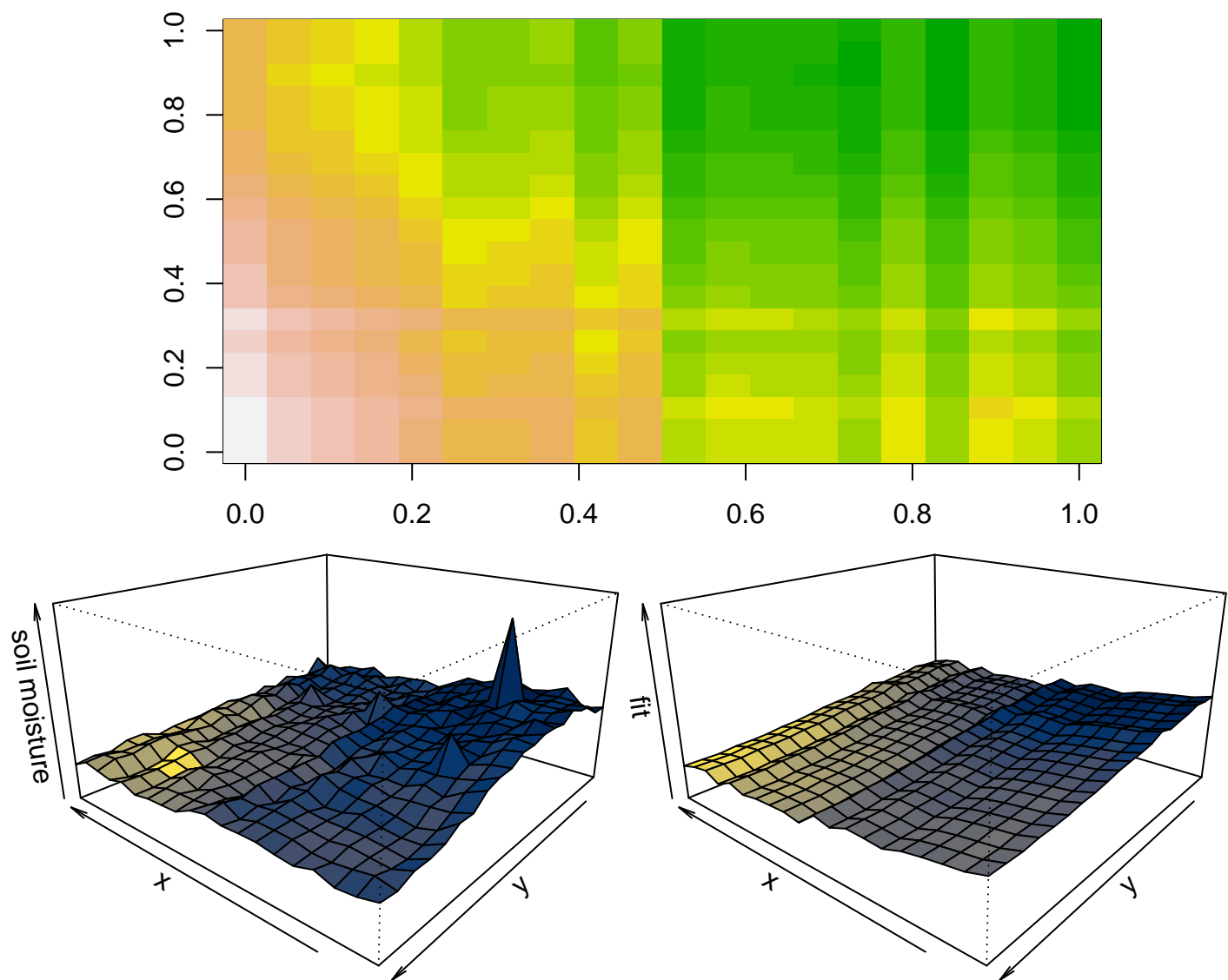


Figure 5: data (left) vs. median polish fit (right)

Median polish seems more robust to the presence of the moisture outlier and does a more reasonable job of modeling the expected layout of the land with moisture projected on top of it. Comparing the median polish fit (right) and the original data (left), in Figure 5, it is clear it does a better job than OLS. but not by much. It captures some local discrepancies across the field, while representing the same overall trend (towards that same corner of high moisture) as OLS.

```

## R version 3.6.2 (2019-12-12)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS Sierra 10.12.6
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods    base
##
## other attached packages:
## [1] zeallot_0.1.0      broom_0.7.0      e1071_1.7-3      assertive_0.3-6
## [5] ggthemes_4.2.0     viridis_0.5.1    viridisLite_0.3.0 here_0.1
## [9] forcats_0.5.0      stringr_1.4.0    dplyr_1.0.2      purrr_0.3.4
## [13] readr_1.3.1        tidyr_1.1.2      tibble_3.0.3     ggplot2_3.3.2
## [17] tidyverse_1.3.0
##
## loaded via a namespace (and not attached):
## [1] httr_1.4.2          jsonlite_1.7.0
## [3] assertive.sets_0.0-3 modelr_0.1.8
## [5] assertthat_0.2.1    assertive.data_0.0-3
## [7] highr_0.8           blob_1.2.1
## [9] cellranger_1.1.0    yaml_2.2.1
## [11] pillar_1.4.6        backports_1.1.9
## [13] glue_1.4.2          assertive.data.uk_0.0-2
## [15] assertive.matrices_0.0-2 digest_0.6.25
## [17] assertive.types_0.0-3 rvest_0.3.6
## [19] colorspace_1.4-1    htmltools_0.5.0
## [21] pkgconfig_2.0.3     assertive.properties_0.0-4
## [23] assertive.data.us_0.0-2 assertive.reflection_0.0-5
## [25] haven_2.3.1         scales_1.1.1
## [27] farver_2.0.3        generics_0.0.2
## [29] ellipsis_0.3.1      withr_2.2.0
## [31] assertive.code_0.0-3 cli_2.0.2
## [33] magrittr_1.5        crayon_1.3.4
## [35] readxl_1.3.1        assertive.strings_0.0-3
## [37] evaluate_0.14       fs_1.5.0
## [39] fansi_0.4.1         assertive.numbers_0.0-2
## [41] xml2_1.3.2          class_7.3-17
## [43] tools_3.6.2         hms_0.5.3
## [45] lifecycle_0.2.0     assertive.files_0.0-2
## [47] munsell_0.5.0       reprex_0.3.0
## [49] compiler_3.6.2      rlang_0.4.7
## [51] grid_3.6.2          rstudioapi_0.11
## [53] assertive.models_0.0-2 assertive.base_0.0-7
## [55] labeling_0.3         rmarkdown_2.3
## [57] gtable_0.3.0        codetools_0.2-16
## [59] DBI_1.1.0           assertive.dates_0.0-3
## [61] R6_2.4.1            gridExtra_2.3
## [63] lubridate_1.7.9     knitr_1.29
## [65] rprojroot_1.3-2     stringi_1.4.6
## [67] Rcpp_1.0.5          vctrs_0.3.4
## [69] dbplyr_1.4.4        tidyselect_1.1.0
## [71] xfun_0.16

```