Buffalo First Analysis

Rachael Caelie (Rocky) Aikens

3/22/2019

```r
read_data <- function(i){
  filename <- paste("../data/nsim_1000/angle_sigma1_results_",i,"_10_1000", sep = "")
  dat <- read.csv(filename) %>%
    mutate(rho = i/10)
  return(dat)
}

dat <- lapply(1:10, read_data) %>% bind_rows
```

Set Up

We compare the performance of propensity score matching, Mahalanobis distance matching, and Buffalo Matching (described in the previous section) on simulated data, varying the dimensionality of the problem, the fixed treatment to control ratio during matching, and the correlation between the true propensity and prognostic score. The generative model for all of our simulations is the following:

$$
\begin{aligned}
X_i &\sim_{iid} \text{Normal}(0, I_p),\\
T_i &\sim_{iid} \text{Bernoulli}\left(\frac{1}{1+\exp(-\phi(X_i))}\right),\\
Y_i &= \tau T_i + \Psi(X_i) + \epsilon_i,\\
\epsilon_i &\sim_{iid} N(0, \sigma^2),
\end{aligned}
$$

where the true propensity and prognositic scores are given by the linear combinations

$$
\begin{aligned}
\phi(X_i) &= X_{i1}/3 - 4,\\
\Psi(X_i) &= \rho X_{i1} + \sqrt{(1-\rho^2)}X_{i2},
\end{aligned}
$$

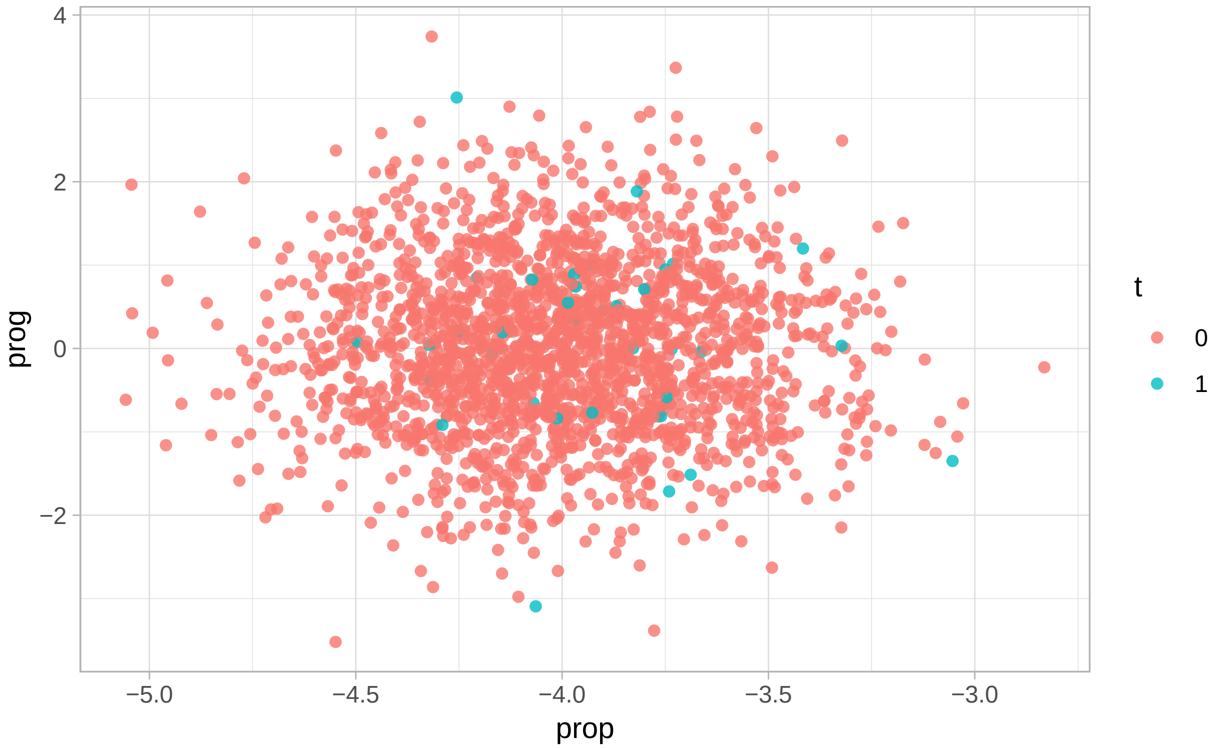
so that $\mathrm{Cor}(\phi(X_i), \Psi(X_i)) \propto \rho$. The propensity score formula was chosen such that there were approximately 100 treated observations in each dataset (But there aren't actually 100 treated in each dataset???). We consider $p = 10$, $\rho = 0, 0.1, \ldots, 0.9, 1.0$, and $k = 1, \ldots, 10$. Each simulation consisted of a dataset of size $n = 2000$ and was repeated $N = 1000$ times. We fix the treatment effect to be constant with $\tau = 1$ and the noise to be $\sigma = 1$. For a given matching, we estimate ATT and design sensitivity $\tilde{\Gamma}$ using the permutation t-statistic from the package `sensitivtymv`

Replication of Dylan's Plots

Basic Visualization

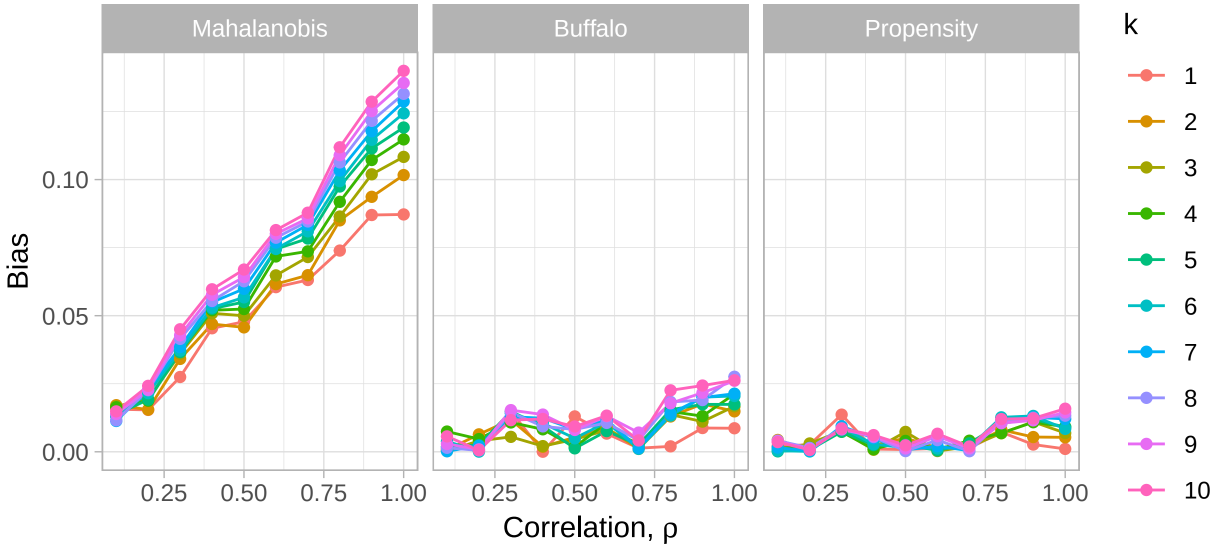
Here, I assume propensity and prognostic information are uncorrelated $\rho = 0$.

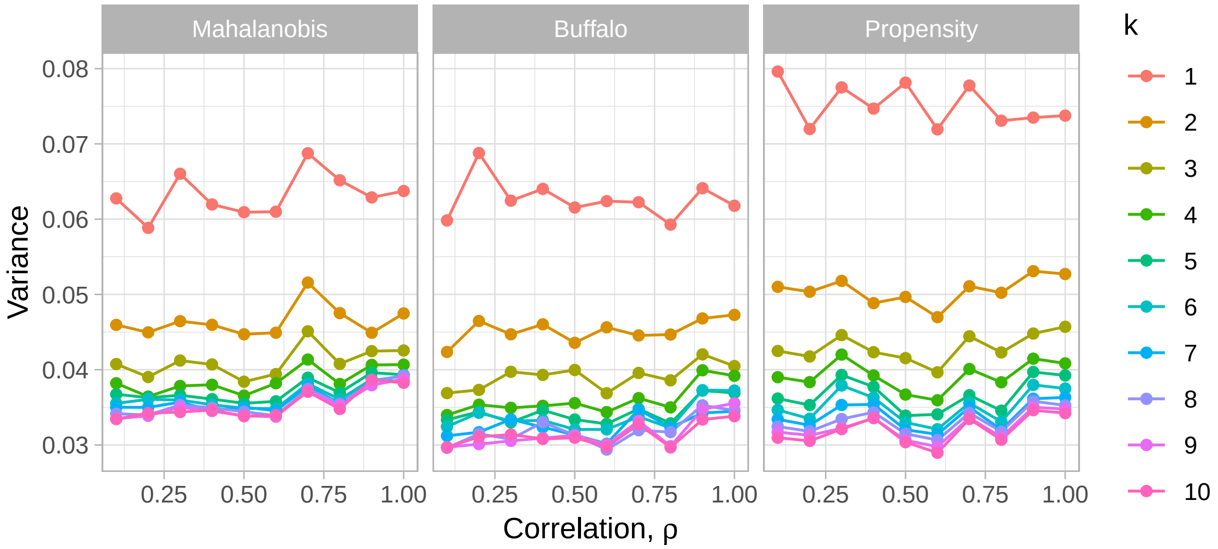
```
#simulate data
df <- generate_data(N = 2000, p = 10, true_mu = "X1/3-4", rho = 0, sigma = 1)
k = 1
prop_model = formula(t ~ . - mu - y)
prog_model = formula(y ~ . - mu - t)

# build propensity score
propensity <- glm(prop_model, family = binomial(), data = df)

prop_match <- pairmatch(propensity, controls = k, df)

# 1:2 mahalanobis matching to select data to use for prognostic model
mahal_dist <- match_on(prop_model, method = "mahalanobis", data = df)
mahal_match <- pairmatch(mahal_dist, controls = 2, df)

buff_match <- prognostic_match(df, propensity, mahal_match, prog_model, k)

# mahalanobis match
m_match <- pairmatch(mahal_dist, controls = k, df)

plt_data <- df %>% mutate(prog = X2, prop = X1/3 - 4, t = as.factor(t)) %>% select(c(t, prog, prop))
ggplot(data = plt_data, aes( x = prop, y = prog, group = t, color = t)) + geom_point(alpha = 0.8)
```

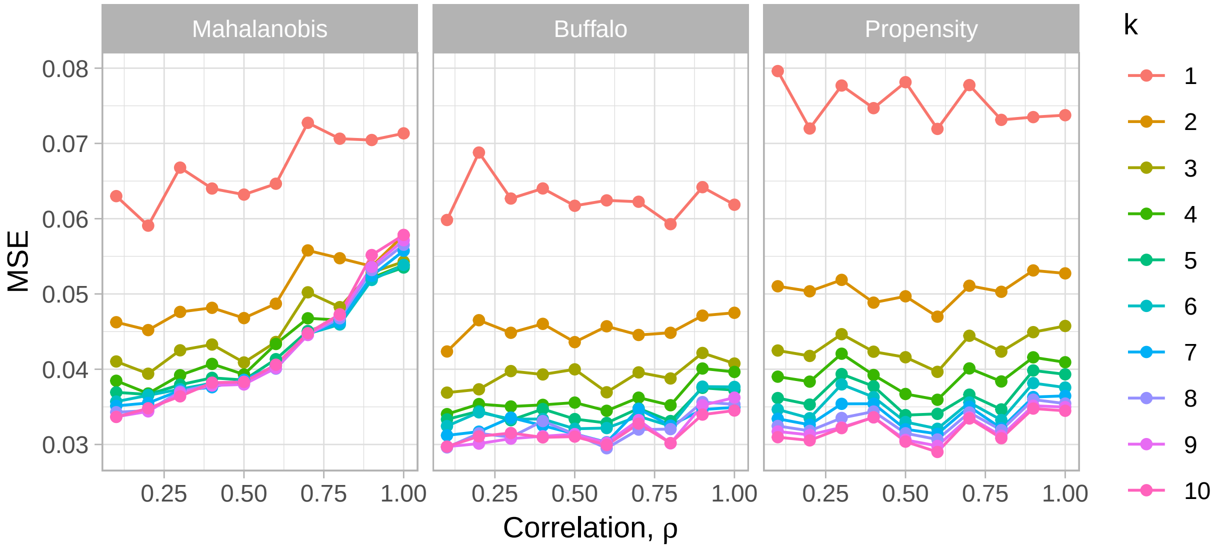


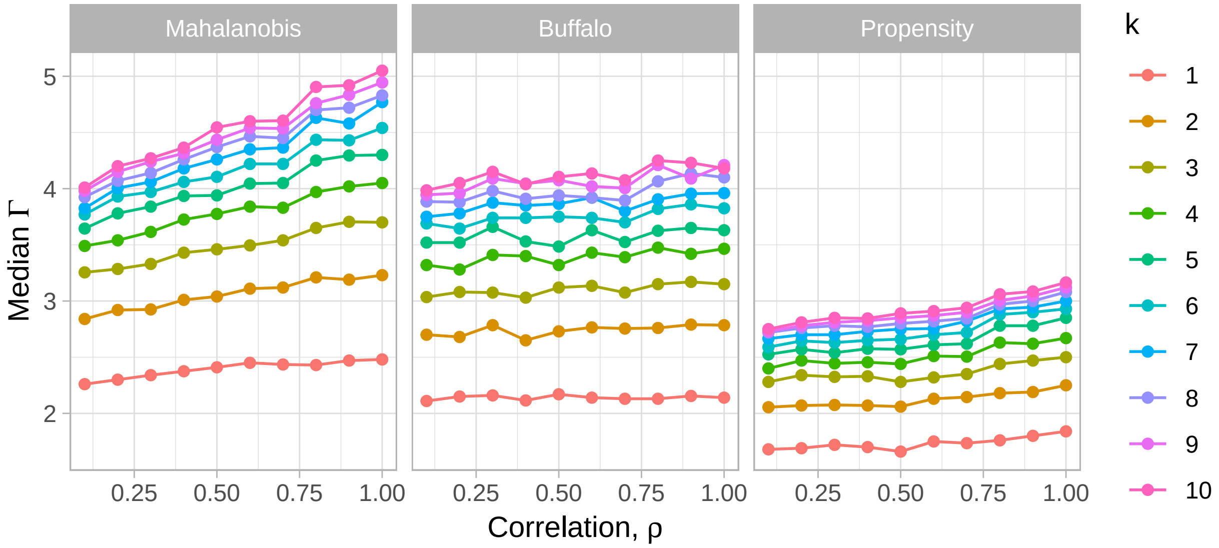
Performance for 1:1 to 1:10 matching as correllation of propensity and prognostic score is modified

```
true_tau <- 1

dat <- mutate(dat,
              squared_err = (estimate-true_tau)**2,
              k = as.factor(k))

plt_data <- dat %>%
  group_by(method, k, rho) %>%
  summarize(Bias = abs(mean(estimate) - true_tau),
            median_gamma = median(gamma),
            Variance = var(estimate),
            MSE = Bias^2 + Variance) %>%
  ungroup() %>%
  mutate(method = recode(method, propensity = "Propensity",
                         mahalanobis = "Mahalanobis",
                         prognostic = "Buffalo"))
```

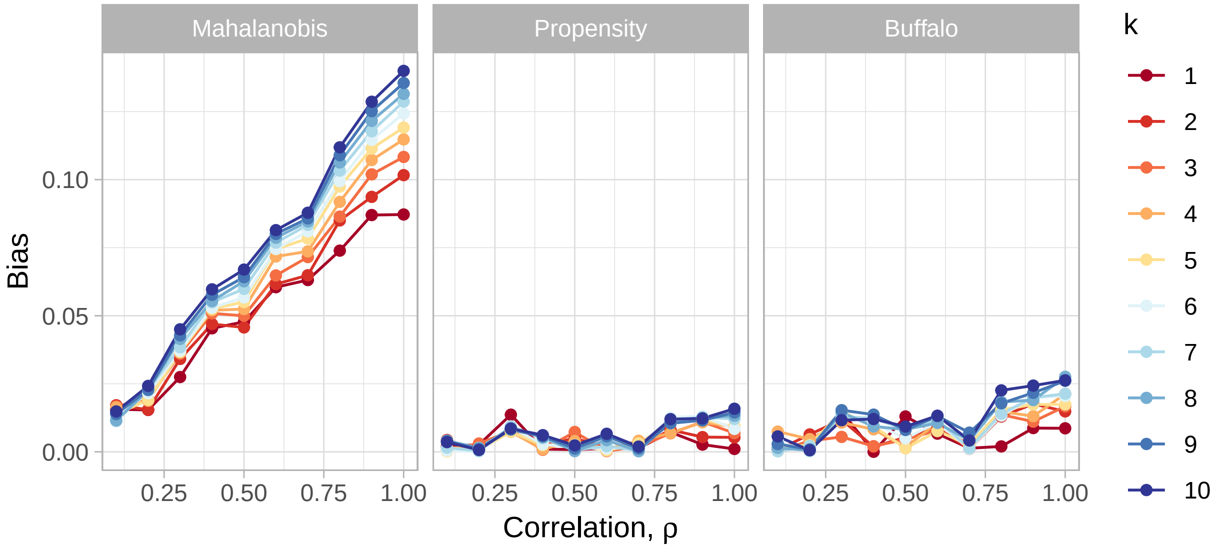


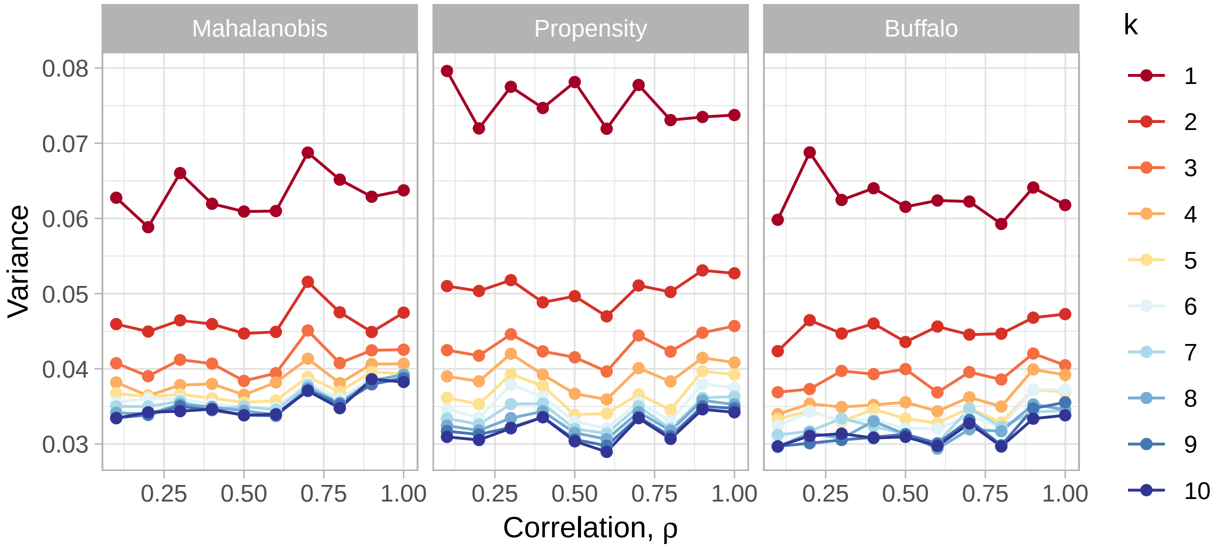

Mahalanobis
Buffalo
Propensity
Variance
0.08
0.07
0.06
0.05
0.04
0.03
0.25
0.50
0.75
1.00
Correlation, ρ
k
1
2
3
4
5
6
7
8
9
10

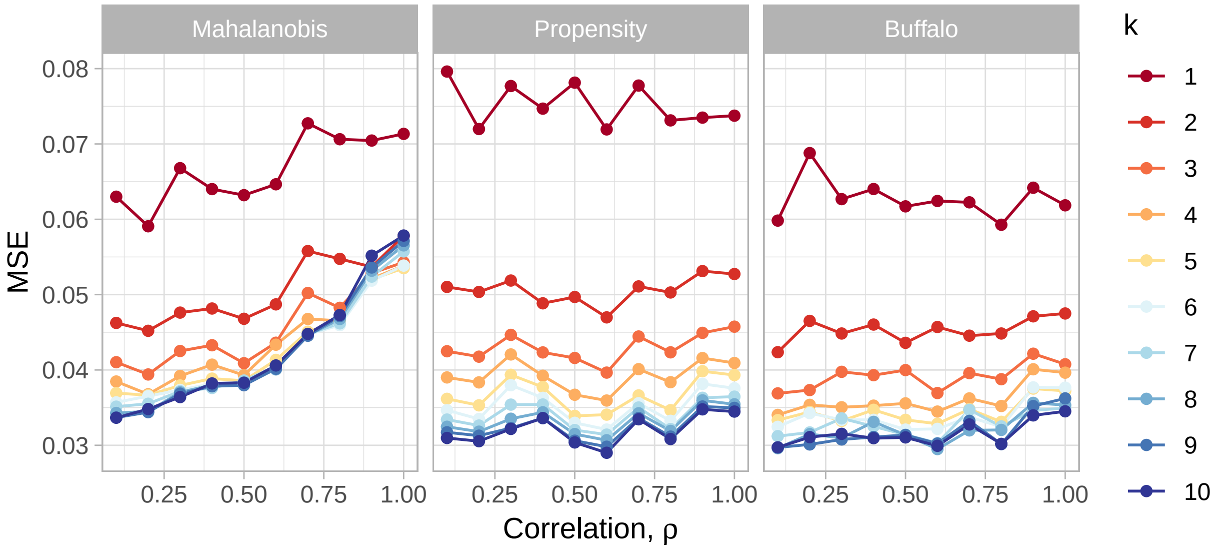

Mahalanobis
Buffalo
Propensity
MSE
0.08
0.07
0.06
0.05
0.04
0.03
0.25
0.50
0.75
1.00
Correlation, ρ
k
1
2
3
4
5
6
7
8
9
10

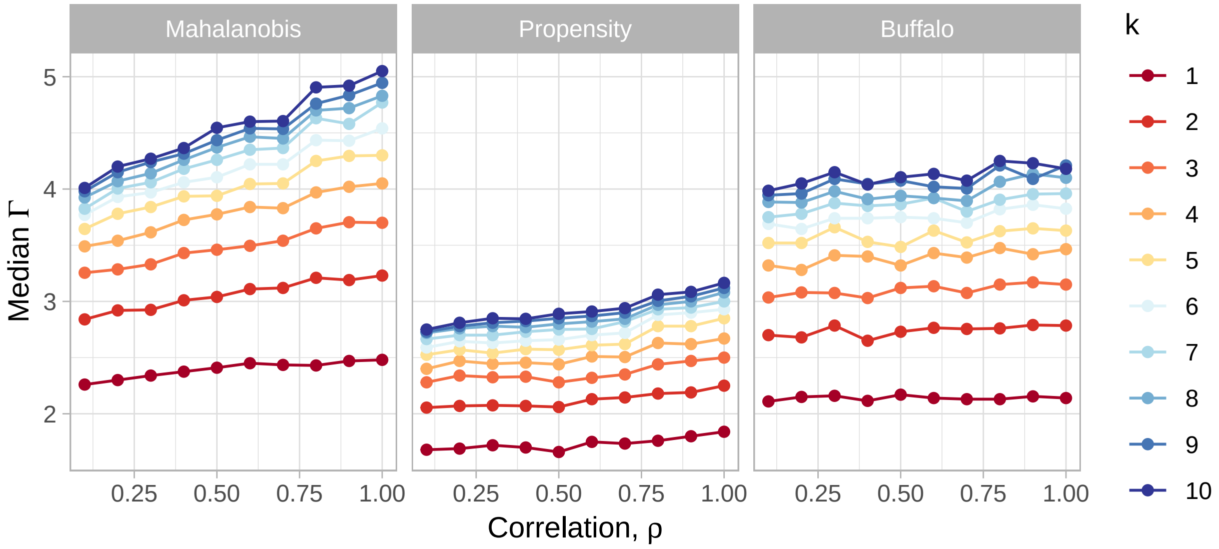

Mahalanobis
Buffalo
Propensity
Median Γ
5
4
3
2
0.25
0.50
0.75
1.00
Correlation, ρ
k
1
2
3
4
5
6
7
8
9
10

Rocky Versions






Mahalanobis
Propensity
Buffalo
MSE
0.08
0.07
0.06
0.05
0.04
0.03
0.25
0.50
0.75
1.00
Correlation, ρ
k
1
2
3
4
5
6
7
8
9
10


Mahalanobis
Propensity
Buffalo
Median Γ
5
4
3
2
0.25
0.50
0.75
1.00
Correlation, ρ
k
1
2
3
4
5
6
7
8
9
10